



# Vinst och förlust

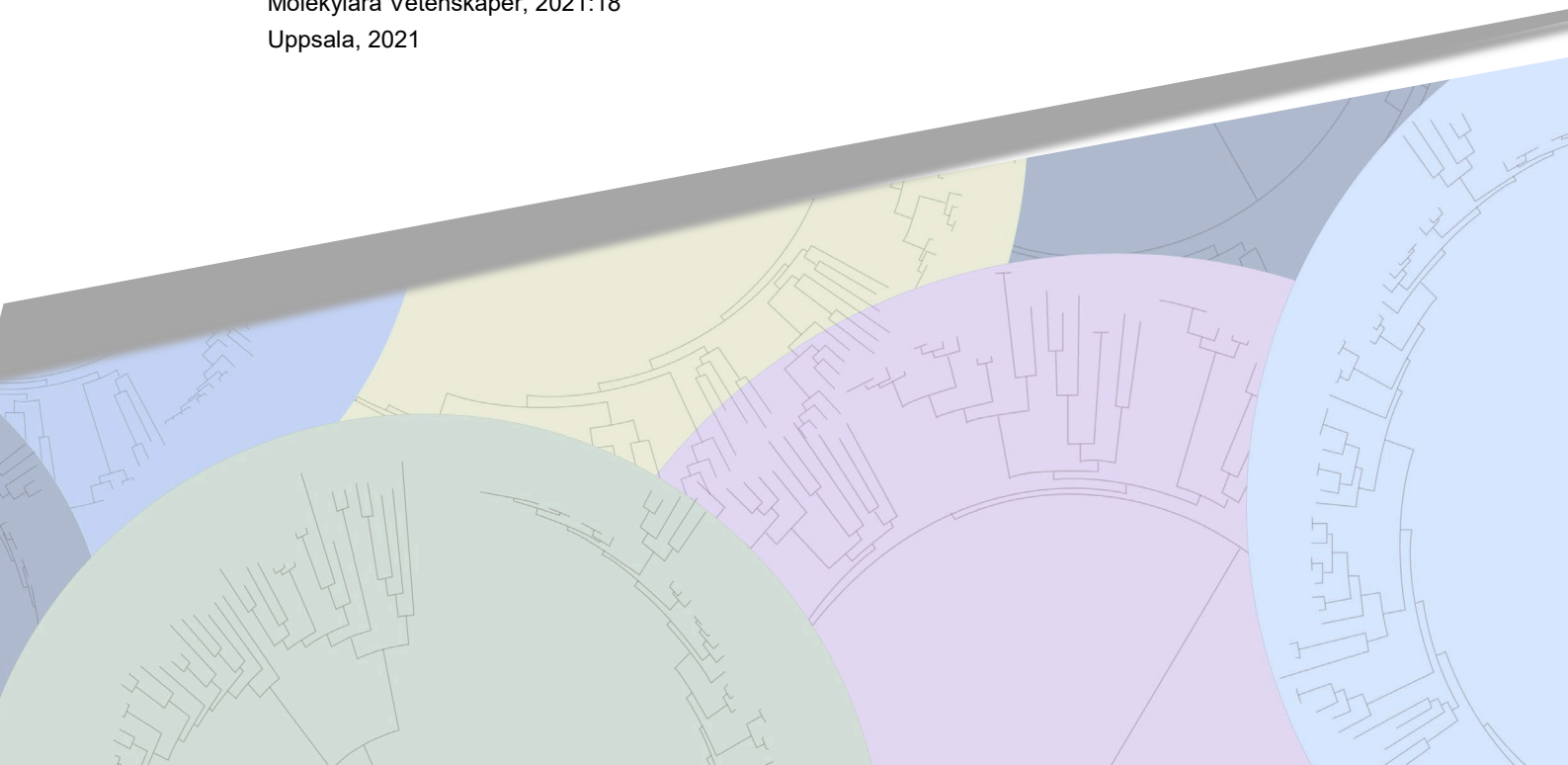
– en fylogenetisk studie av kloroperoxidas-liknande enzym hos knoppjästsvampar

---

*Gain and loss – a phylogenetic study of putative chloroperoxidases found within budding yeasts*

Beatrice Bylén

Examensarbete/Självständigt arbete • (15 hp)  
Sveriges lantbruksuniversitet, SLU  
Institutionen för molekylära vetenskaper  
Agronomprogrammet, mark/växt  
Molekylära Vetenskaper, 2021:18  
Uppsala, 2021



# Vinst och förlust – en fylogenetisk studie av kloroperoxidas-liknande enzym hos jästsvampar

*Gain and loss – a phylogenetic study of putative chloroperoxidases found within budding yeasts*

Beatrice Bylén

**Handledare:** Tomas Linder, SLU, Institutionen för Molekylära Vetenskaper  
**Examinator:** Peter Bozhkov, SLU, Institutionen för Molekylära Vetenskaper

**Omfattning:** 15 hp  
**Nivå och fördjupning:** Grundnivå, G2E  
**Kurstitel:** Självständigt arbete i biologi  
**Kurskod:** EX0894  
**Program/utbildning:** Agronomprogrammet, mark/växt  
**Kursansvarig inst.:** Institutionen för Vatten och Miljö

**Utgivningsort:** Uppsala  
**Utgivningsår:** 2021  
**Omslagsbild:** Beatrice Bylén  
**Serietitel:** Molekylära Vetenskaper  
**Delnummer i serien:** 2021:18

**Nyckelord:** fylogenetik, sekvenshomologi, Saccharomycotina, knoppjästsvamp, HGT, genförlust,  $\alpha/\beta$ -hydrolas, kloroperoxidas

**Sveriges lantbruksuniversitet**

Fakulteten naturresurser och jordbruksvetenskap  
Institutionen för Molekylära Vetenskaper

## Publicering och arkivering

Godkända självständiga arbeten (examensarbeten) vid SLU publiceras elektroniskt. Som student äger du upphovsrätten till ditt arbete och behöver godkänna publiceringen. Om du kryssar i **JA**, så kommer fulltexten (pdf-filen) och metadata bli synliga och sökbara på internet. Om du kryssar i **NEJ**, kommer endast metadata och sammanfattning bli synliga och sökbara. Fulltexten kommer dock i samband med att dokumentet laddas upp arkiveras digitalt.

Om ni är fler än en person som skrivit arbetet så gäller krysset för alla författare, ni behöver alltså vara överens. Läs om SLU:s publiceringsavtal här: <https://www.slu.se/site/bibliotek/publicera-och-analysera/registrera-och-publicera/avtal-for-publicering/>.

☒ JA, jag/vi ger härmed min/vår tillåtelse till att föreliggande arbete publiceras enligt SLU:s avtal om överlåtelse av rätt att publicera verk.

☐ NEJ, jag/vi ger inte min/vår tillåtelse att publicera fulltexten av föreliggande arbete. Arbetet laddas dock upp för arkivering och metadata och sammanfattning blir synliga och sökbara.

## Sammanfattning

Knoppjästsvampar är en typ av svampar med encellig morfologi. Knoppjästsvampar utgör Saccharomycotina, en understam till Ascomycota och innefattar mer än 1000 idag beskrivna arter, bland annat bakjästen *Saccharomyces cerevisiae*. Trots att *S. cerevisiae* är en av de mest studerade eukaryota organismerna förekommer ett stort antal andra arter av knoppjästsvampar som först nyligen sekvenserats och där ytterst få funktioner hos gener är kartlagda. Gener utan klarlagda funktioner brukar kallas "orphan genes".

Som grund för den fylogenetiska analysen i detta projekt ligger en icke-karaktäriserad genfamilj hos knoppjästsvampar som kodar för ett förmodat kloroperoxidas. Syftet var att försöka upprätta en hypotes om enzymets funktion baserat på de ekologiska förhållanden där arter av knoppjästsvampar med denna typ av gen återfinns. Sökningar gjordes efter homologa sekvenser till den valda genen, både inom gruppen knoppjästsvampar och inom andra taxonomiska grupper. De påträffade homologerna låg till grund för konstruktionen av ett fylogenetiskt träd som sedan användes som underlag för en hypotes om relationen mellan de kloroperoxidas-liknande gener som påträffades.

Undersökningen resulterade i ett fylogenetiskt träd där endast en del av alla arter av knoppjästsvampar bildade en monofyletisk grupp. Cirka två tredjedelar av de påträffade sekvenserna bildade en större monofyletisk grupp, medan de andra sekvenserna bildade mindre grupper som visade närmre släktskap med bakteriella kloroperoxidaser. Denna polyfyletiska fördelning av sekvenser från knoppjästsvamparna tyder på att nära evolutionärt släktskap finns mellan dessa och de bakteriella sekvenserna i analysen. Detta släktskap kan bero på en eller flera fall av horisontell genöverföring från bakterie till knoppjästsvamp.

*Nyckelord:* fylogenetik, sekvenshomologi, Saccharomycotina, knoppjästsvamp, HGT, genförlust,  $\alpha/\beta$ -hydrolas, kloroperoxidas

## Abstract

Budding yeasts are a group of predominantly unicellular fungi that together form the subphylum Saccharomycotina of phylum Ascomycota. Subphylum Saccharomycotina consists of over 1000 described species, which include the common baker's yeast *Saccharomyces cerevisiae*. *S. cerevisiae* is one of the most studied eukaryotic organisms but until recently only a few other species of budding yeasts have had their genomes sequenced. Despite the recent increase in sequenced budding yeast genomes, genome annotation has lagged behind resulting in many so-called "orphan genes", genes without a described function, which have yet to be examined.

The focus of this project was a non-characterized gene family within budding yeasts that encode a putative chloroperoxidase. The purpose was to construct a hypothesis concerning enzyme function based on ecological or metabolic properties of the species where chloroperoxidase-like homologues were found. Searches for homologues for the chosen gene were performed both within Saccharomycotina and in other taxonomic groups. The found homologues formed a dataset which was used to construct a phylogenetic tree. This tree functioned as support for a hypothesis about possible relationships between species where homologous were found.

The analysis generated a phylogenetic tree where only a part of the sequences from budding yeasts constituted a monophyletic group. Two thirds of sequences found in budding yeasts formed a larger monophyletic group, while the remaining third formed smaller groups that exhibited closer relationships with bacterial chloroperoxidase than with the yeast sequences. This polyphyletic distribution of budding yeast sequences indicates a closer evolutionary relationship to bacterial sequences. This relationship could be due to one or several instances of horizontal gene transfer from bacteria to yeast.

**Keywords:** phylogenetics, sequence homology, Saccharomycotina, yeast, HGT, gene loss,  $\alpha/\beta$ -hydrolase, chloroperoxidase



# Innehållsförteckning

|  |           |
|--|-----------|
| <b>Tabellförteckning.....</b>                                  | <b>8</b>  |
| <b>Figurförteckning .....</b>                                  | <b>9</b>  |
| <b>Förkortningar .....</b>                                     | <b>10</b> |
| <b>1. Inledning .....</b>                                      | <b>11</b> |
| 1.1. Bakgrund.....   | 11        |
| 1.2. Syfte och frågeställning .....                            | 14        |
| <b>2. Metod.....</b>   | <b>15</b> |
| 2.1. Insamling av data .....                                   | 15        |
| 2.2. Sekvensjämförelse och preliminära fylogenetiska träd..... | 16        |
| 2.3. Konstruktion av slutgiltigt fylogenetiskt träd.....       | 17        |
| <b>3. Resultat.....</b>  | <b>18</b> |
| 3.1. Resultat från sökningar .....                             | 18        |
| 3.2. Fylogenetisk analys.....                                  | 27        |
| <b>4. Slutsatser och diskussion .....</b>                      | <b>29</b> |
| 4.1. Genförlust .....  | 30        |
| 4.2. Horisontell genöverföring.....                            | 31        |
| 4.3. Kombination av event .....                                | 34        |
| 4.4. Enzymatisk funktion .....                                 | 34        |
| 4.5. Förslag på vidare forskning.....                          | 35        |
| 4.6. Slutsats .....  | 35        |
| <b>Referenser.....</b>   | <b>37</b> |
| <b>Tack.....</b>   | <b>39</b> |
| <b>Bilaga 1 .....</b>  | <b>40</b> |

## Tabellförteckning

|  |    |
|--|----|
| Tabell 1. Resultat för sökningar mot databaserna x och x inom understammen Saccharomycotina.....                               | 19 |
| Tabell 2. Resultat för sökningen mot databas x inom understammen Saccharomycotina.....   | 20 |
| Tabell 3. Resultat för homologisökningar utanför understammen Saccharomycotina.....  | 25 |
| Tabell 4. Mest signifikant träff för de sekvenser som valts ut som representativa för sin gren i det fylogenetiska trädet..... | 26 |



## Figurförteckning

- Figur 1. Kladogram över förhållandet mellan kladerna i understammen Saccharomycotina. Avståndet mellan noder och klader är inte exakt, utan syftar till att ge en översikt förhållandet mellan dem. Uppdelningen av kladerna är hämtade från Shen. et al 2018.....12
- Figur 2. Taxonomisk fördelning av påträffade homologer till genen som kodar för det förmodade  $\alpha/\beta$ -hydrolaset hos *D. hansenii* (GenBank proteinaccessionsnummer XP\_456694). Släkten inom vilka homologer påträffats är markerade med rektanglar. Uppdelningen av kladerna är hämtad från Shen et al. (2018).....21
- Figur 3. Fördelning av arter inom de grupperingar av knoppjästsvampar där homologer till sekvensen *D. hansenii* (XP\_456694) påträffats.....22
- Figur 4. Orotat fylogenetiskt träd. Bootstrapvärden  $> 0.5$  är angivna vid respektive nod. Monofyletiska grupper bestående av sekvenser från knoppjästsvampar samt sekvenser från enstaka knoppjästsvampar är markerade med färg. Sekvenser från Pezizomycotina är markerade med grått.....28
- Figur 5. Schematisk bild över var ett potentiellt HGT-eventet kan skett som gjort att kloroperoxidas-liknande sekvenser erhållits av en gemensam förfader till understammarna Pezizomycotina och Saccharomycotina. Det fylogenetiska förhållandet mellan understammarna är hämtat från Cuomo & Birren (2010) .....33

# Förkortningar

Artträd – ett fylogenetiskt träd som illustrerar arters förhållande till varandra

BLAST, Basic Local Alignment Search Tool- en sökmotor för att jämföra likheter hos nukleotid- eller proteinsekvenser. Fungerar också som ett verktyg för att söka upp homologer

Bootstrapping – en itereringsprocess i vilken ett stort antal fylogenetiska träd konstrueras varvid det slutgiltiga är konstruerat utifrån medelvärde från de tidigare

HGT-event, horizontal gene transfer event – Ett tillfälle där genöverföring mellan organismer som inte beror på sexuell överföring skett

Genomannotation- identifiering av gener i ett genom

Genträd – ett fylogenetiskt träd som illustrerar släktskap för en specifik gen

Klad - En monofyletisk grupp, en grupp av grenar som alla härstammar och utgår från en viss nod

WGD, whole genome duplication – ett event där hela en organisms genom dupliceras

ORF, Open Reading Frame – Hypotetiska protein-kodande nukleotidsekvenser som börjar och slutar med ett start- respektive slutkodon utan mellanliggande icke-kodande sekvens

Sekvensjämförelse - jämförelse av två eller fler nukleotid- eller proteinsekvenser för att identifiera bevarade positioner

Sekvenshomologi – biologiska sekvenser med gemensamt evolutionärt ursprung

Utgrupp – en extern nod i ett fylogenetiskt träd som inte ingår i en systergrupp. Kan användas som referens för att skapa rotade fylogenetiska träd

# 1. Inledning

## 1.1. Bakgrund

Jästsvampar är ett samlingsbegrepp för de svampar som kan föröka sig asexuellt som enskilda celler genom avknoppning (Blackwell 2017). Jästsvampar utgör alltså ingen monofyletisk grupp, utan jäst är en levnadsform som förekommer utspritt i riket Fungi och återfinns både inom stammarna Basidiomycota och Ascomycota.

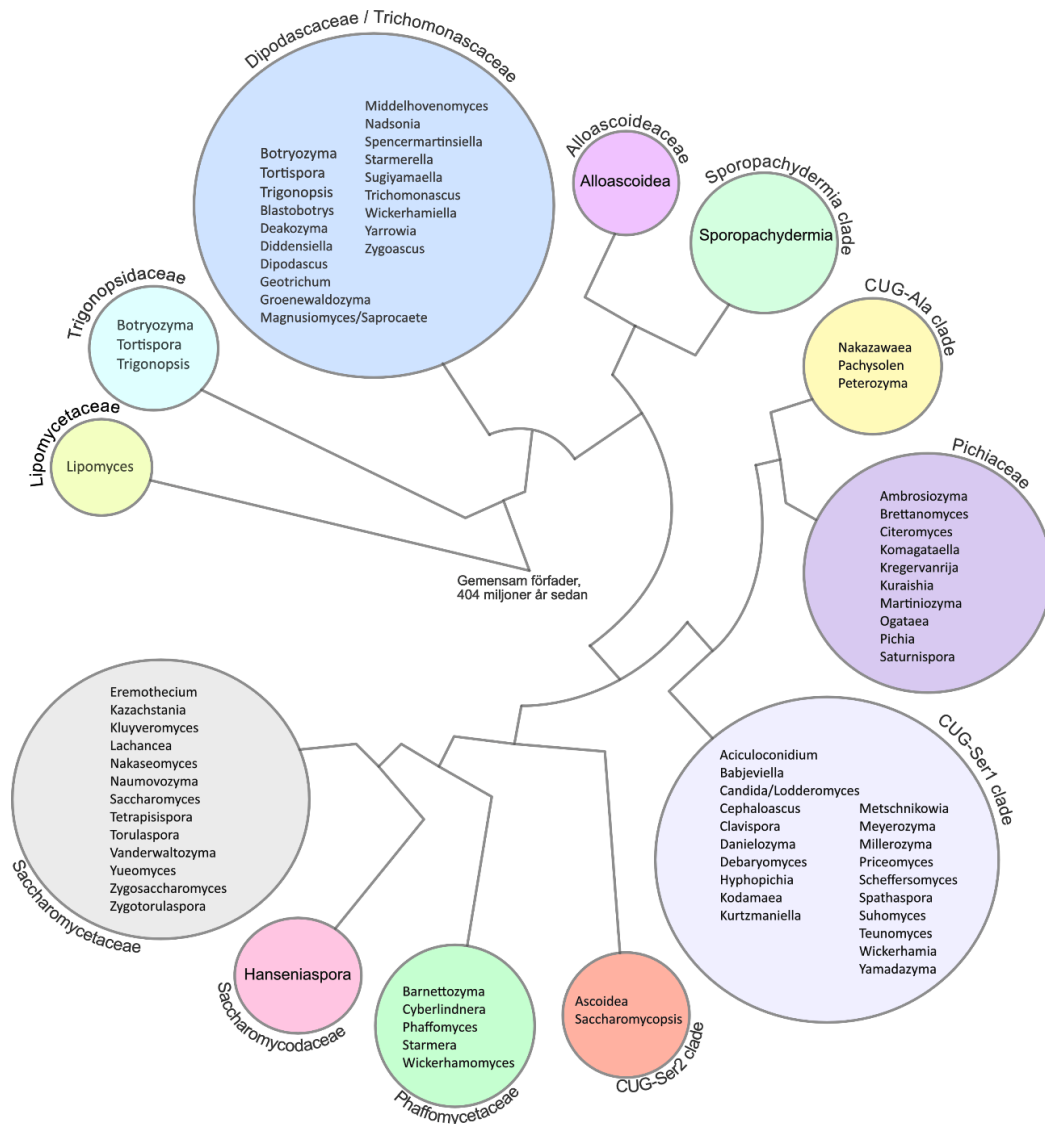
När man i vardagligt tal pratar om jäst åsyftas oftast bakjästen *Saccharomyces cerevisiae* som ingår i understammen Saccharomycotina (inom stam Ascomycota) (Hittinger et al. 2015). I denna grupp ingår mer än 1000 i dagsläget beskrivna arter. Gruppen Saccharomycotina brukar kallas för "true yeasts" eller knoppjästsvampar då de flesta arter endast reproducerar sig genom avknoppning (Blackwell & Spatafora 2004).

Tack vare sin diversitet och specialisering vad gäller ekologi förekommer knoppjästsvampar i en mängd olika miljöer och habitat (Kurtzman 2011). De har hittats på alla kontinenter och i alla biom. Bakom jästernas framgång i så diversa miljöer ligger en mängd olika metaboliska processer som tillåter olika arter att leva på specifika substrat (Shen et al. 2018).

Många jästsvampar är saprofyter, men många arter är också förknippade med olika typer av symbioser (Kurtzman 2011). Ett exempel på symbios är de knoppjästsvampar som lever mutualistiskt med insekter (Blackwell 2017). Jästsvamparna förser insekterna med steroler och vitaminer, medan jästsvamparna troligtvis drar nytta av insekternas tarmar som en stabil miljö. Insekterna fungerar också som vektorer till nya habitat med nya näringskällor för jästsvamparna.

2018 sammanställdes och analyserades 332 genom från knoppjästsvampar för att konstruera ett fylogenetiskt träd (Shen et. al). Tidigare har andra fylogenetiska analyser gjorts för att systematiskt organisera knoppjästsvampar, men hittills har ingen så övergripande analys gjorts som samlat så många arter. Analysen resulterade i ett träd med 12 klader och en tidslinje över knoppjästsvamparnas

utveckling från en gemensam förfader för 404.2 miljoner år sedan, den så kallade urknoppjätten. Undersökningen omfattade och kartlade också HGT-events under knoppjästsvamparnas utveckling. Trots att många jästgenom blivit sekvenserade på senare tid är det fortfarande få som blivit annoterade (Proux-Wéra et al. 2012).



Figur 1. Kladogram över förhållandet mellan kladerna i understammen Saccharomycotina. Avståndet mellan noder och klader är inte exakt, utan syftar till att ge en översikt förhållandet mellan dem. Uppdelningen av kladerna är hämtade från Shen. et al 2018

Kodonet CUG translateras i de flesta organismer till aminosyran leucin. Undantag finns dock, Krassowski et al. (2018) visade att vissa knoppjästsvampar inte translaterar CUG till leucin, utan att vissa av arterna translaterar CUG till alanin och vissa till serin. Denna variation ligger till grund för namngivningen av kladerna så som klassifikationen enligt Shen et al. (2018) där tre klader kallas CUG-Ala, CUG-ser1 respektive CUG-ser2.

”Orphan genes” är ett uttryck som använts för gener som blivit annoterade men där det hypotetiska proteinet de kodar för ännu inte har en känd och bekräftad funktion (Hanson et al. 2009). Karaktärisering och bestämning av proteiners funktioner på laborativ väg förmår inte ha samma snabba takt som storskalig sekvensering av genom. Därför förblir många gener orphan genes, vilket är fallet i flera knoppjästgenom. Generna kan få hypotetiska funktioner som baserats på konserverade domäner, men dessa är bara spekulativa. Dessa orphan genes utgör ett utforskat landskap med stor potential.

Superfamiljen  $\alpha/\beta$ -hydrolas innefattar en rad enzymer med olika funktioner så som lipaser, proteaser och haloperoxidaser (Mindrebo et al. 2016). Denna typ av enzym återfinns i alla domäner av cellulära organismer och verkar inom många essentiella metaboliska processer. Gemensamt för enzymerna i denna superfamilj är en katalytisk triad bestående tre aktiva centrum där aminosyran histidin alltid är konserverad (Dimitriou et al. 2017). Haloperoxidaser är en typ av  $\alpha/\beta$ -hydrolaser, som i närvaro av halider och peroxid ( $H_2O_2$ ) katalyserar halogenisering av organiska substanser (Hofmann et al. 1998). Kloroperoxidas är en särskild typ av haloperoxidas, som använder klorid som den halid som oxideras i reaktionen.

Kloroperoxidas har tidigare studerats i bakterien *Rhodococcus erythropolis* (De Schrijver et al. 1997). Bakterien återfinns i jord och kan bryta ned en rad olika organiska substanser, bland annat har *R. erythropolis* uppmärksamats för sin förmåga att bryta ned herbicider innehållande thiokarbamater med hjälp av kloroperoxidas. Sekvenshomologer till bakteriella haloperoxidaser finns hos vissa knoppjästsvampar, men deras biologiska funktion har ännu inte studerats hos svampar.

Fylogenetiska träd är ett sätt att använda sig av sekvensdata för att visualisera och bilda hypoteser om släktskap (Choudhuri 2014). Detta sätt att undersöka evolutionärt släktskap är användbart för att undersöka orphan genes då en gens ursprung kan ge indikationer om dess funktion. Träden kan till exempel ge ledtrådar om vad som skett under en arts utveckling i form av till exempel HGT-events eller helgenomsdupliceringar. Viktigt att komma ihåg är att ett fylogenetiskt träd inte uppvisar ett definitivt evolutionärt släktskap, utan endast är en grafisk representation av en hypotes om släktskap

Fylogenetiska träd upprättas genom att DNA- eller proteinsekvenser sekvenseras och jämförs (Choudhuri 2014). I trädet representerar de interna noderna den senaste gemensamma förfadern för den art eller grupp av organismer som de externa noderna representerar (Baldauf 2003).

Ett träd kan vara rotat eller icke-rotat beroende på vad som undersöks (Baldauf 2003). Trädets rot bör utgöras av den sista universella gemensamma förfadern, går det inte att avgöra vilken denne är bör trädet lämnas icke-rotat.

## 1.2. Syfte och frågeställning

Syftet med projektet var att undersöka om några mönster kunde uppvisas mellan de arter av knoppjästsvampar hos vilka homologer till kloroperoxidas hittats. Projektet består av en bioinformatisk studie där sekvenser jämförs. Jämförelsen ligger till grund för konstruktionen av ett fylogenetiskt träd.

Frågeställningar:

- Går det att urskilja ett mönster mellan de arter av knoppjästsvampar där den kloroperoxidas-liknande genen hittats?
- Går det att uppskatta en hypotetisk funktion för enzymet baserat på dessa knoppjästsvampars ekologiska eller fysiologiska profiler?

## 2. Metod

För att kunna konstruera ett fylogenetiskt träd behövs homologa sekvenser att ställa upp och jämföra mot varandra (Baldauf 2003). För att samla in data användes BLAST (Altschul et al. 1990). BLAST (Basic Local Alignment Search Tool) kan användas för att jämföra likheter hos nukleotid- eller proteinsekvenser samt fungerar som ett verktyg för att söka upp homologer. Resultaten från sökningarna redovisas efter lägst e-värde till högst, där e-värdet representerar sannolikheten att träffen beror på slumpen i stället för faktisk likhet mellan sekvenserna. Ett cutoff-värde är det högsta e-värde för vilka träffarna bedöms vara signifikanta. Om inget annat anges sattes gränsvärdet för sökningarna till  $10^{-50}$ .

Alla sökningar utgick från en kloroperoxidas-liknande sekvens från knoppjästsvampen *Debaryomyces hansenii* (GenBank protein accession XP\_456694). Under insamlingen av data konstruerades med jämna mellanrum preliminära fylogenetiska träd med hjälp av neighbor-joiningmetoden. Dessa preliminära träd användes som underlag för att ta bort sekvenser som bedömdes otillförlitliga eller icke-relevanta för analysen. Denna process av att testa datasetet i form av provisoriska fylogenetiska träd upprepades tills resultatet bedömdes som tillfredsställande.

### 2.1. Insamling av data

Grunden till datasetet består av sökningar i BLASTP mot understammen *Saccharomycotina* (NCBI:txid147537) i databaserna Non-redundant protein sequences och Reference sequence (refseq) proteins. Motsvarande sökning gjordes också med protein-mot-nukleotid-sökning (TBLASTN) mot Whole genome shotgun contig (wgs)-databasen. Eftersom sekvenserna i wgs contig-databasen är icke-annoterade nukleotidsekvenser användes NCBI's ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>) för att identifiera kompletta proteinsekvenser för hypotetiska kloroperoxidas-liknande gener inom wgs-databasen. För translationen valdes korresponderande translationstabell för respektive art. De translaterade sekvenserna jämfördes sedan i TBLASTN mot den

genomiska sekvensen för att få fram exakta genomiska koordinater för den kodande DNA-sekvensen.

För att utöka datasetet bortom knoppjästsvampar utfördes även sökningar mot andra områden understammar och mot andra organismtyper som beskrivs i resultatsektionen.

## 2.2. Sekvensjämförelse och preliminära fylogenetiska träd

För att identifiera och justera konserverade sekvenspositioner inom samtliga sekvenser för användning inom fylogenetisk analys användes flarsekvensjämförelseprogrammet MAFFT (Kato et al. 2005; <https://mafft.cbrc.jp/alignment/server/index.html>) med G-INS-i som jämförelsemetod. För att trimma bort icke-konserverad sekvens från de uppställda sekvenserna användes Gblocks (Castresana 2000; [http://molevol.ibmb.csic.es/Gblocks\\_server/](http://molevol.ibmb.csic.es/Gblocks_server/)). Sekvenser som bedömdes för korta eller som på andra sätt påverkade resultaten negativt gallrades bort, och uppställning och trimning upprepades. Som inställningar för trimningen valdes selektionsalternativen "Allow smaller final blocks" samt "Allow less strict flanking positions". För att göra preliminära fylogenetiska träd av de trimmade sekvenserna användes MAFFT phylogenetics (Kato et al. 2005; <https://mafft.cbrc.jp/alignment/server/phylogeny.html>) För konstruktionen av trädet valdes neighbor joining-metoden med konserverade positioner. För "Heterogeneity among sites" valdes "estimate". För att få ett säkert träd valdes bootstrapping med 1000 iterationer.

Sekvenser från arterna *D. hansenii* (GenBank protein accession XP\_456694), *Wickerhamiella sorbophila* (GenBank protein accession XP\_024665284), *Candida intermedia* (GenBank protein accession SGZ48255), *Meyerozyma guilliermondii* (GenBank protein accession SGZ48255), *Methanosarcina acetivorans* (GenBank protein accession WP\_011021030) samt *Methanoculleus thermophilus* (GenBank protein accession WP\_066955142) valdes ut som representanter för sina grupperingar i det fylogenetiska trädet så som det såg ut vid detta steg i processen. Sökningar gjordes i BLASTP mot var och en av sekvenserna utan avgränsning gällande taxonomi. Det resultat med högst likhet för respektive art valdes ut och togs med i datasetet.



### 2.3. Konstruktion av slutgiltigt fylogenetiskt träd

För att konstruera ett slutgiltigt fylogenetiskt träd användes en metod för att konstruera trädet baserat på maximal sannolikhet. Mjukvaran PhyML v. 3 (<http://www.atgc-montpellier.fr/phyml/>) valdes. Sekvenserna linjerades och trimmades på samma sätt som tidigare. Sekvenserna formaterades sedan till phylip-format genom LIRMMs formatkonverterare ([http://phylogeny.lirmm.fr/phylo\\_cgi/data\\_converter.cgi](http://phylogeny.lirmm.fr/phylo_cgi/data_converter.cgi)). För val av modell till konstruktion av trädet användes PhyMLs funktion att automatiskt välja modell, "automatic model selection by sms" (Lefort et al. 2017). Modellen som valdes var LG + G + I . Datatypen sattes till aminosyror.

Det konstruerade trädet visualiserades sedan i MEGA-X v.10.0.05 (Kumar et al. 2018) och formaterades med hjälp Affinity Designer v. 1.9 (<https://affinity.serif.com/en-us/>). Ett icke-rotat träd valdes då ingen tydlig utgrupp fanns att använda som rot.

En formaterad flersekvensjämförelse av de sekvenserna som valts ut som slutgiltigt dataset, se bilaga 1, gjordes med Boxshade (Hofmann & Baron 070521). Inställning för hur stor andel av positionernas identiteter som måste stämma överens för att de ska bli skuggade (fraction of sequences) sattes till 0.7.

## 3. Resultat

### 3.1. Resultat från sökningar

Sökningarna utgick ifrån en kloroperoxidas-likande proteinsekvens (GenBank protein accession XP\_456694) från knoppjästsvampen *D. hansenii* eftersom denna sekvens uppvisade signifikant likhet med bakteriella proteinsekvenser som kodar för karakteriserade kloroperoxidas. Till exempel visar det bakteriella kloroperoxidaset från *Rhodococcus erythropolis* (GenBank protein accession AAC45285) 54.64 % sekvensidentitet med den utvalda sekvensen från *D. hansenii*, e-värde  $9 \times 10^{-111}$ .

För sökningarna inom Saccharomycotina mot databaserna Non-redundant protein sequences och Reference proteins noterades tydliga ”hopp” där e-värdet ökade kraftigt mellan två sökresultat. Ett sådant hopp indikerar att de enzym med e-värden högre än innan hoppet inte har samma funktion som det enzym som söktes med. Dessa hopp låg till grund för avgränsningen vad gällde vilka resultat som bedömdes signifikanta. För sökningen mot databasen Non-redundant protein sequence sattes avgränsningen till  $6 \times 10^{-94}$ , och för sökning mot Reference proteins sattes avgränsning till  $2 \times 10^{-94}$ .

Tabell 1. Resultat för sökningar mot databaserna Non-redundant protein sequence och Reference proteins inom understammen Saccharomycotina

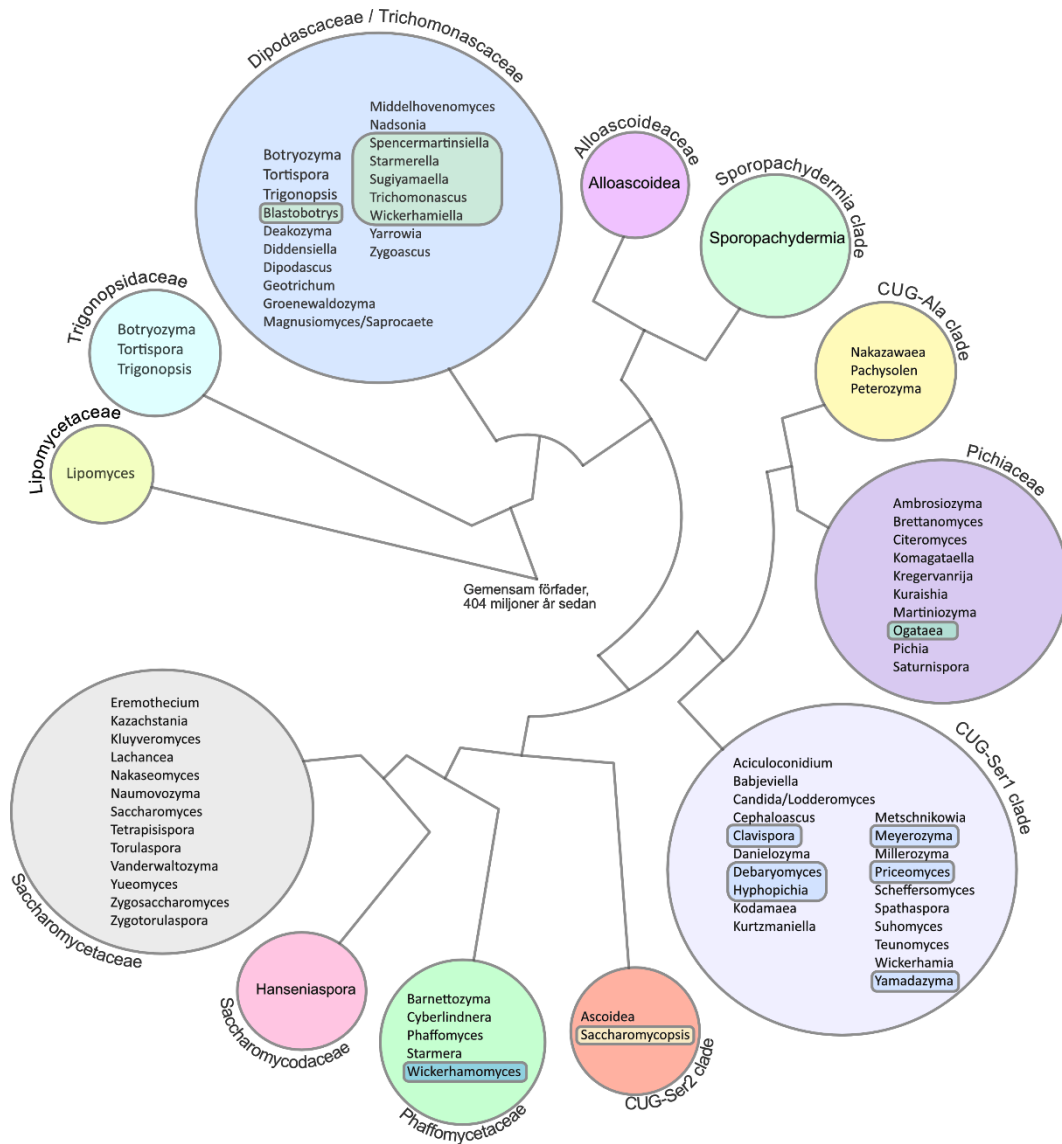
| Art                              | Accessionsnummer |
|----------------------------------|------------------|
| <i>Debaryomyces hansenii</i>     | XP_456694        |
| <i>Debaryomyces fabryi</i>       | XP_015465771     |
| <i>Wickerhamomyces anomalus</i>  | XP_019040844     |
| <i>Meyerozyma guilliermondii</i> | SGZ48255         |
| <i>Candida auris</i>             | XP_028892780     |
| <i>Candida haemuloni</i>         | XP_025342312     |
| <i>Hyphopichia burtonii</i>      | XP_020074028     |
| <i>Candida duobushaemulonis</i>  | XP_025336863     |
| <i>Candida pseudohaemulonii</i>  | XP_024714720     |
| <i>Ogataea polymorpha</i>        | XP_018211057     |
| <i>Wickerhamiella sorbophila</i> | XP_024665284     |
| <i>Candida intermedia</i>        | SGZ48255         |
| <i>Meyerozyma sp. JA9</i>        | RLV88139         |
| <i>Trichomonascus ciferrii</i>   | KAA8917649       |

Även för motsvarande sökning mot Whole genome shotgun contig-databasen noterades ett hopp och resultat med e-values lägre än  $1 \times 10^{-83}$  togs med i datasetet.

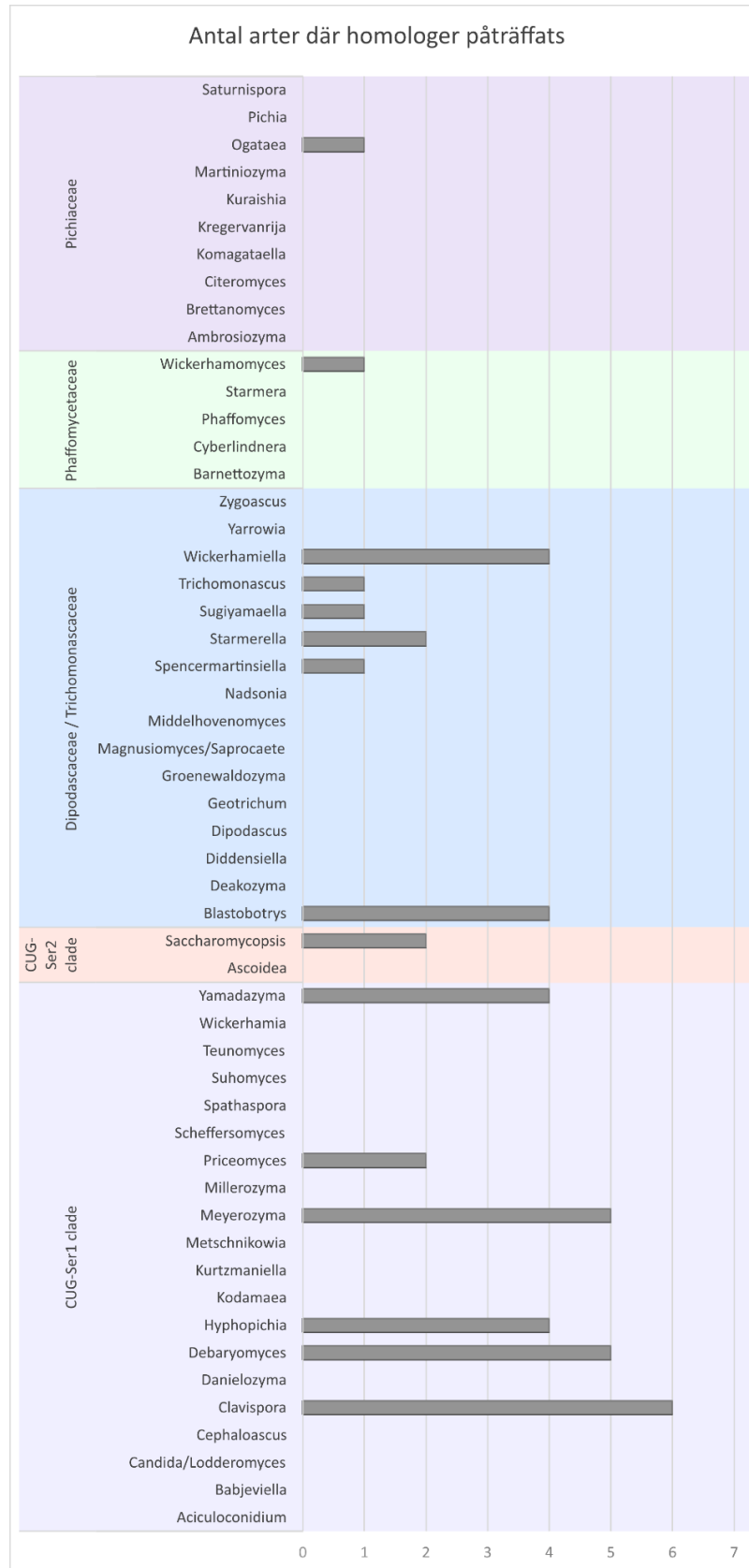
Tabell 2. Resultat för sökningen mot databas Whole genome shotgun contig inom understammen Saccharomycotina. Genetisk kod syftar till den kod som användes för translation till hypotetisk proteinsekvens. Kod 1 motsvarar standardkod, kod 12 motsvarar alternative yeast nuclear-kod.

| Art                                 | Genomiskt        |        | Genetisk |                 |
|-------------------------------------|------------------|--------|----------|-----------------|
|                                     | accessionsnummer | Sträng | kod      | Koordinater     |
| <i>Debaryomyces subglobosus</i>     | PPOW01000005     | -      | 12.      | 480313-481140   |
| <i>Debaryomyces prosopidis</i>      | PPOQ01000002     | +      | 12.      | 281038-281865   |
| <i>Debaryomyces maramus</i>         | PPOS02000008     | +      | 12.      | 5814-6638       |
| <i>Priceomyces carsonii</i>         | PPMM02000016     | +      | 12.      | 268593-269420   |
| <i>Priceomyces medius</i>           | PPIF01000002     | -      | 12.      | 259929-260762   |
| <i>Candida ascalaphidarum</i>       | PPPA02000035     | -      | 12.      | 16497-17318     |
| <i>Candida tammaniensis</i>         | PPOY01000002     | +      | 12.      | 385745-386563   |
| <i>Hyphopichia heimii</i>           | PPMQ02000004     | -      | 12.      | 277578-278405   |
| <i>Meyerozyma athensensis</i>       | PPIU01000016     | -      | 12.      | 16797-17624     |
| <i>Candida carpophila</i>           | BCGK01000006     | -      | 12.      | 727032-727847   |
| <i>Candida gotoi</i>                | PPNP01000040     | +      | 12.      | 11108-11935     |
| <i>Meyerozyma caribbica</i>         | PPNE02000017     | -      | 12.      | 50517-51332     |
| <i>Candida gorgasii</i>             | PPPC02000016     | +      | 12.      | 10298-11119     |
| <i>Candida blattae</i>              | PPMS02000016     | -      | 12.      | 99943-100761    |
| <i>Hyphopichia homilentoma</i>      | BCGB01000006     | +      | 12.      | 1029320-1030147 |
| <i>Saccharomycopsis malanga</i>     | BCGJ01000003     | -      | 12.      | 1122211-1123041 |
| <i>Spencermartinsiella europaea</i> | PPIE02000011     | +      | 1.       | 265886-266713   |
| <i>Blastobotrys mokoensis</i>       | PPJM02000026     | +      | 1.       | 164274-165101   |
| <i>Blastobotrys serpentis</i>       | PPJG01000078     | +      | 1.       | 9800-10627      |
| <i>Blastobotrys raffinosa</i>       | PPJH01000011     | -      | 1.       | 67740-68567     |
| <i>Wickerhamiella domercqiae</i>    | NREC01000001     | +      | 1.       | 13475-14308     |
| <i>Wickerhamiella infanticola</i>   | NREI01000009     | +      | 1.       | 69348-70172     |
| <i>Saccharomycopsis fodiens</i>     | JNFV01000005     | -      | 12.      | 197159-197977   |
| <i>Candida aaseri</i>               | LKAN01000038     | -      | 12.      | 11967-12794     |
| <i>Sugiyamaella xylanicola</i>      | MQSX02000034     | +      | 1.       | 105782-106609   |
| <i>Starmerella ratchasimensis</i>   | NRDV01000063     | +      | 1.       | 20985-21812     |
| <i>Wickerhamiella occidentalis</i>  | NREE01000014     | +      | 1.       | 179936-180766   |
| <i>Starmerella sp. 'tilneyi'</i>    | NREA01000015     | -      | 1.       | 165797-166753   |
| <i>Blastobotrys adenivorans</i>     | CBZY010000006    | +      | 1.       | 1591928-1592911 |

För dessa tre sökningar inom Saccharomycotina påträffades homologer hos totalt 43 arter. Dessa arter är sett till taxonomin väldigt spritt fördelade. Homologerna påträffades i 14 släkten fördelade över 4 klader, se figur 2. De arter där homologer påträffades är framför allt koncentrerade till släkten inom kladerna CUG-Ser1 samt till Dipodascaceae/Trichomonascaceae, se figur 3.



Figur 2. Taxonomisk fördelning av påträffade homologer till genen som kodar för det förmodade  $\alpha/\beta$ -hydrolaset hos *D. hansenii* (GenBank proteinaccessionsnummer XP\_456694). Släkten inom vilka homologer påträffats är markerade med rektanglar. Uppdelningen av kladerna är hämtad från Shen et al. (2018)



Figur 3. Fördelning av arter inom de grupperingar av knoppjästsvampar där homologer till sekvensen *D. hansenii* (XP\_456694) påträffats.

För att utöka datasetet utanför Saccharomycotina gjordes flera sökningar för att undersöka om homologer finns hos andra grupper inom riket Fungi (NCBI:txid4751). I och med att projektets fokus är knoppjästsvampar valdes snäva cutoff-värden för sökningar utanför Saccharomycotina där träffarna blev väldigt många. Detta för att inte ta med alla träffar i datasetet då det skulle bli för stort och innehålla för många sekvenser som inte kommer från knoppjästsvampar. I stället valdes i dessa fall de mest signifikanta träffarna ut och togs med som en representation för alla resultaten i just den sökningen.

Sökningar inom riket Fungi gjordes på de två andra understammarna till Ascomycota förutom Saccharomycotina: Pezizomycotina (NCBI:txid147538) och Taphrinomycotina (NCBI:txid451866).

För att precisera sökningen inom Pezizomycotina gjordes vidare sökningar i klasserna Eurotiomycetes (NCBI:txid147545) och Pezizomycetes (NCBI:txid147549). Dessa klasser valdes ut då de innefattar arterna *Aspergillus nidulans* respektive *Tuber melanosporum*, två välstuderade arter. Skulle homologer hittas i välstuderade arter är chansen högre att funktionerna hos dessa proteiner blivit bestämda.

För dessa sökningar hittades ett stort antal homologer inom underklassen Pezizomycotina och inom klassen Eurotiomycetes. Därför valdes höga cutoff-värden,  $6 \times 10^{-104}$  respektive  $5 \times 10^{-98}$ . Ingen träff hittades för arten *A. nidulans* eller *T. melanosporum*. För sökningen inom Pezizomycetes hittades inga träffar under gränsvärdet. Inte heller några resultat påträffades för sökningen mot understammen Taphrinomycotina.

För att vidare undersöka riktet Fungi gjordes en sökning mot stammen Basidiomycota (NCBI:txid5204). Denna sökning gav endast två träffar med signifikant resultat. När en sökning ger sådana udda resultat med enstaka träffar kan kontaminering misstänkas. För att kontrollera detta gjordes BLAST sökningar efter närmast liknande organism, där träffarna för de båda sekvenserna blev andra Basidiomycotasvampar. Detta i kombination med att generna båda befinner sig på kromosomer samt att de båda är långt större än 10 000 baspar ledde till bedömningen att de inte är kontaminanter trots att de var de enda två resultaten inom avgränsningarna.

En sökning mot hela riket Fungi där Basidiomycota (NCBI:txid5204) och Ascomycota (NCBI:txid4890) exkluderades gjordes. Motsvarande sökning gjordes mot domänen Eukaryota (NCBI:txid2759) där riket Fungi exkluderades.

För dessa sökningar hittade inga resultat under gränsvärdet som inte bedömdes vara kontaminanter. De resultat som påträffades med låga e-värden kom från enstaka

spridda organismer och när sökningar på dessa gjordes mot alla databaser var gemensamt för de högsta träffarna att de kom från bakteriesekvenser. Detta tyder på att sekvenseringarna blivit kontaminerade och att en bakterie sekvenserats istället för den eukaryota organismen.

Sökningar gjordes mot hela domänen Archaea (NCBI:txid2157). På grund av ett stort antal resultat sattes ett snävt värde för avgränsning till  $3 \times 10^{-101}$  och de mest signifikanta resultaten togs med som en representation för domänen. Totalt adderades sex arkéesekvenser till datasetet.

En sökning utan exkluderande parametrar gjordes också mot Protein Data Bank-databasen med förhoppningen att kunna addera sekvenser för protein med tillskrivna funktioner till datasetet. Resultat från denna databas skulle kunna hjälpa till att understödja en teori om enzymets funktion. Värdet för avgränsning sattes här  $1 \times 10^{-79}$  vilket resulterade i fem resultat under gränsvärdet. Gemensamt för dessa resultat är att de är bakteriesekvenser, två av dem från familjerna Streptomycetaceae, två från Pseudomonadaceae och en från Burkholderiaceae.

För att inkludera ytterligare bekräftade kloroperoxidas-sekvenser i jämförelsen hämtades sekvenser från artiklar där de enzymatiska funktionerna hos bakteriella kloroperoxidaser har beskrivits (De Schrijver et al. 1997); (De Mot et al. 2003). Dessa sekvenser jämfördes mot sekvensen från *D. hansenii* genom verktyget för att jämföra två eller fler sekvenser i BLASTP. De sekvenserna med e-värden inom gränsvärdet likhet togs med i datasetet, totalt tre sekvenser.

Svampen *Leptoxylum fumago* producerar ett kloroperoxidas (Huster et al. 2016). Detta kloroperoxidas tillhör en annan genfamilj av peroxidaser, men inkluderades sökningen togs med i analysen för jämförelse. En sökning gjordes mot *L. fumago* (NCBI:txid5474), utan resultat innanför gränsvärdet. En sökning gjordes därför mot Capnodiaceae (NCBI:txid62133), familjen där släktet *Leptoxylum* ingår. Gränsvärdet sattes till  $2 \times 10^{-104}$ , med ett resultat innanför gränsvärdet.



Tabell 3. Resultat för homologisökningar utanför understammen *Saccharomycotina*

| Sökning                                   | Resultat                            |                  |
|---|-------------------------------------|------------------|
|   | Art                                 | Accessionsnummer |
| Pezizomycotina<br>(NCBI:txid147538)       | <i>Fonsecaea monophora</i>          | XP_022509128     |
|   | <i>Fonsecaea multimorphosa</i>      | XP_016631922     |
|   | <i>Mollisia scopiformis</i>         | XP_018073521     |
|   | <i>Exophiala oligosperma</i>        | XP_016257131     |
|   | <i>Cyphellophora europaea</i>       | XP_008713467     |
|   | <i>Phialemoniopsis curvata</i>      | XP_030998524     |
|   | <i>Phialophora attinorum</i>        | XP_018002380     |
|   | <i>Scytalidium lignicola</i>        | RFU31724         |
|   | <i>Lachnellula occidentalis</i>     | TVY36247         |
|   | <i>Rhizodiscina lignyota</i>        | KAF2092632       |
|   | <i>Helotiales</i> sp.               | KAE8447949       |
|   | <i>Cadophora malorum</i>            | KAG4415059       |
|   | <i>Cadophora</i> sp.                | PVH73381         |
|   | <i>Neonectria ditissima</i>         | KPM44821         |
|   | <i>Cladophialophora bantiana</i>    | XP_016623295     |
| Eurotiomycetes<br>(NCBI:txid147545)       | <i>Fonsecaea erecta</i>             | XP_018695805     |
|   | <i>Exophiala aquamarina</i>         | XP_013265785     |
|   | <i>Fonsecaea pedrosoi</i>           | XP_013288316     |
|   | <i>Cladophialophora immunda</i>     | XP_016250788     |
|   | <i>Cladophialophora psammophila</i> | XP_007750673     |
|   | <i>Mycena galopus</i>               | KAF8145373       |
| Basidiomycota<br>(NCBI:txid5204)          | <i>Mycena venus</i>                 | KAF7344751       |
|   | <i>Methanosarcina acetivorans</i>   | WP_011021030     |
| Archaea<br>(NCBI:txid2157)                | <i>Methanosarcina horonobensis</i>  | WP_048138765     |
|   | <i>Methanosarcina siciliae</i>      | WP_048173570     |
|   | <i>Methanoculleus thermophilus</i>  | WP_066955142     |
|   | <i>Methanosarcina</i> sp. Ant1      | OEU43333         |
|   | <i>Pseudomonas fluorescens</i>      | 1A8S_A           |
| PDB-databas, utan begränsade parametrar   | <i>Streptomyces lividans</i>        | 1A88_A           |
|   | <i>Burkholderia cenocepacia</i>     | 4DGQ_A           |
|   | <i>Pseudomonas putida</i>           | 1ZOI_A           |
| Sekvenser från artiklar om kloroperoxidas | <i>Rhodococcus erythropolis</i>     | AAC45285         |
|   | <i>Burkholderia pyrrocinia</i>      | WP_175964102     |
|   | <i>Synechocystis</i> sp.            | WP_028946588     |
| Capnodiaceae<br>(NCBI:txid62133)          | <i>Polychaeton citri</i>            | KAF2719266       |

Utifrån ett preliminärt fylogenetiskt träd valdes sekvenser från arterna *D. hansenii* (GenBank protein accession XP\_456694), *Wickerhamiella sorbophila* (GenBank protein accession XP\_024665284), *Candida intermedia* (GenBank protein accession SGZ48255), *Meyerozyma guilliermondii* (GenBank protein accession SGZ48255), *Methanosarcina acetivorans* (GenBank protein accession WP\_011021030) samt *Methanoculleus thermophilus* (GenBank protein accession WP\_066955142) ut som representanter för sina grupperingar.

Sökningar gjordes i BLASTP mot var och en av sekvenserna utan avgränsning gällande taxonomi. Det resultat med högst likhet för respektive art valdes ut och togs med i datasetet. Detta gjordes med förhoppning att den högsta träffen skulle kunna indikera något om huruvida sekvensen har sitt ursprung från till exempel en bakterie eller en svamp som inte är en knoppjästsvamp. Samtliga resultat kom från bakterier, se tabell 4.

Tabell 4. Mest signifikant träff för de sekvenser som valts ut som representativa för sin gren i det fylogenetiska trädet

| Representativ sekvens från trädet                | Art                         | Accessionsnummer |
|--|-----------------------------|------------------|
| <i>Debaryomyces hansenii</i> (XP_456694)         | <i>Serratia odorifera</i>   | WP_004964504     |
| <i>Wickerhamiella sorbophila</i> (XP_024665284)  | <i>Novosphingobium rosa</i> | WP_068082913     |
| <i>Candida intermedia</i> (SGZ48255)             | <i>Serratia</i> sp.         | NMC23253         |
| <i>Methanosarcina acetivorans</i> (WP_011021030) | <i>Pedospira parvula</i>    | WP_007415115     |

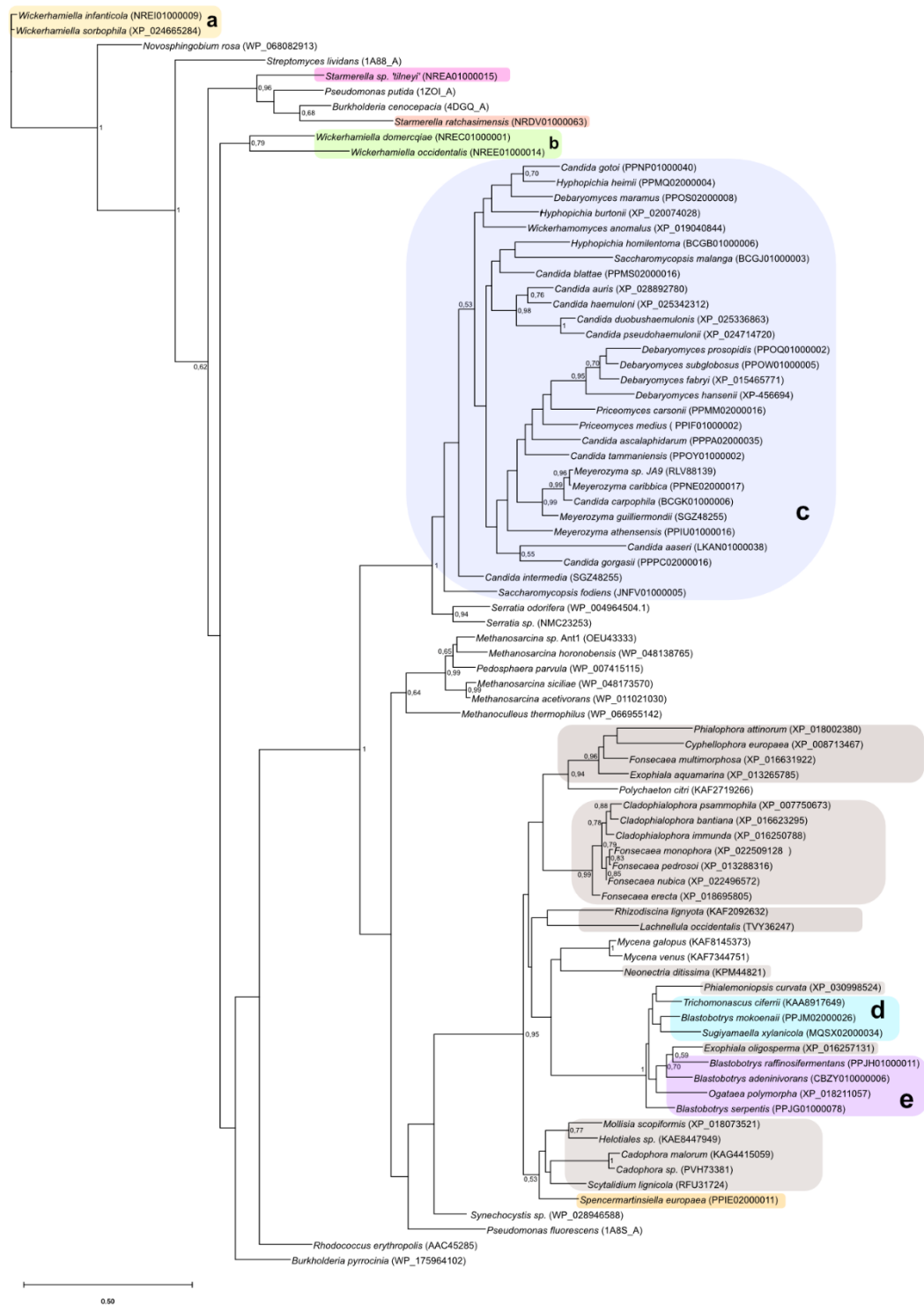
Efter att datasetet sammanställts gjordes den sekvensjämförelse som ligger till grund för konstruktionen av det fylogenetiska trädet. Sekvensjämförelsen resulterade i 259 konserverade positioner som lämpade sig för fylogenetisk analys. Grafisk representation för sekvenslinjeringen ses i bilaga 1.

### 3.2. Katalytisk triad

Som underlag för att undersöka vilka sekvenser i datasetet som uppvisar den katalytiska triad som är typisk för  $\alpha/\beta$ -hydrolas användes sekvensen från *R. erythropolis* (GenBank protein accession AAC45285). Denna sekvens kodar för ett bekräftat kloroperoxidas. De tre aktiva centrumen som utgör den katalytiska triaden befinner sig i sekvensen på position 95, 225 samt 254 och utgörs av aminosyrorna serin, asparagin respektive histidin (Uniprot n.d.). Dessa positioner jämfördes med motsvarande positionerna i flersekvensjämförelsen, se bilaga 1. Jämförelsen visade att samtliga sekvenser i datasetet uppvisade den katalytiska triaden.

### 3.3. Fylogenetisk analys

Det fylogenetiska träd som konstruerats uppvisar fem monofyletiska grupper av knoppjästsvampar, se figur 5. En av dessa är större och omfattade alla sekvenshomologer som påträffats i kladerna CUG-Ser1, CUG-Ser2 och familjen Phaffomycetaceae, grupp c i figuren. De fyra andra monofyletiska grupperna i genträdet består av färre sekvenser och befinner relativt långt ifrån varandra i genträdet. De två sekvenserna från knoppjästsvamparna *Starmerella sp. 'tilneyi'* och *Spencermartinsiella europaea* tillhör i trädet inga monofyletiska grupper. De fyra arterna från *Wickerhamiella*-släktet är i trädet fördelade på två monofyletiska grupper som befinner sig relativt åtskilt.



Figur 4. Orotat fylogenetiskt träd. Bootstrapvärden > 0.5 är angivna vid respektive nod. Monofyletiska grupper bestående av sekvenser från knoppjästsvampar samt sekvenser från enstaka knoppjästsvampar är markerade med färg. Sekvenser från Pezizomycotina är markerade med grått.

## 4. Slutsatser och diskussion

Syftet med detta projekt var att undersöka den taxonomiska förekomsten av ett hypotetiskt kloroperoxidas hos knoppjästsvampar, med förhoppning att kunna uppskatta en preliminär funktion för detta enzym. Förhoppningen var att se ett tydligt samband mellan ekologin hos de arter vars genom innehöll homologer till bakteriella kloroperoxidaser. Det förväntades också att samtliga kloroperoxidas-liknande enzymen hos knoppjästsvampar skulle bilda en monofyletisk grupp på grund av gemensamt evolutionärt ursprung. I stället för en stabil monofyletisk grupp så bildade de kloroperoxidas-likande generna hos knoppjästsvampar ett flertal enskilda förgreningar.

Sett till hela understammen Saccharomycotina finns kloroperoxidas-liknande gener enbart i en handfull släkten, och inom dessa släkten finns inte genen heller inom alla arter. Vissa sekvenser som undersökts i detta projekt är även de enda som påträffats i sin klad.

Diskussionen av dessa resultat kommer att utgå från tre möjliga evolutionära scenarion:

- En homolog till den undersökta genen har funnits i urknoppjästen och sedan tappats någonstans på vägen till de arter som inte uppvisar någon homolog.
- En homolog till genen fanns inte hos urknoppjästen, utan har erhållits av en eller flera förfäder till de arter som visar upp homologer idag.
- En kombination av de två övre scenarion, en kombination av genförlust och erhållning.

## 4.1. Scenario I – Genförlust

Knoppjästsvampar har små genom och genförlust är något som är väl utbrett (Shen et al. 2018). Troligtvis är det denna tendens till reduktiv evolution som drivit utvecklingen framåt och gjort knoppjästsvamparna så metaboliskt diversa. Mönstret att kloroperoxidas-liknande gener påträffats i flera släkten, men i släktena bara av några enstaka arter skulle tyda på att förlust av genen skett. Om så är fallet har genen tappats hos de arter som inte uppvisar homologer, till exempel på grund av specialisering till en miljö där enzymet genen kodar för inte är nödvändigt.

Något som talar för att genen har funnits hos den så kallade urknoppjästen till Saccharomycotina är att homologer påvisats även i understammen Peziziomycotina. Sekvenserna från Peziziomycotina bildar ingen enhetlig fylogenetisk grupp, utan befinner sig precis som knoppjästsvamparna uppdelade i flera fylogenetiska grupper, se figur 4. Om vi utgår från att de kloroperoxidas-liknande generna i Saccharomycotina och Peziziomycotina har gemensamt ursprung och har nedärvts vertikalt bör alltså denna gen funnits i den gemensamma förfadern till de båda understammarna. Eventuellt har en homolog till genen funnits i förfadern till alla tre understammarna, men att genförlust skett under utvecklingen av Taphrinomycotina och att det är därför genen saknar representation i den understammen.

De flesta arter av knoppjästsvampar där homologer till den kloroperoxidas-linkande gener hittats ingår i kladerna Dipodasceaeae/Trichomonascaseae och CUG-Ser1. De enda som inte ingår i dessa två klader är *Ogataea polymorpha*, *Wickerhamomyces anomalus* samt *Saccharomycopsis malanga* och *Saccharomycopsis fodiens*. Trots att dessa fyra arter rent taxonomiskt befinner sig på vitt skilda ställen inom Saccharomycotina har de alla placerat sig i den större monofyletiska gruppen c i genträdet, se figur 4. Detta skulle kunna tyda på att de är de enda enstaka arterna som har kvar en sekvenshomolog efter att andra arter i deras klader har förlorat genen.

## 4.2. Scenario II – Horisontell genöverföring

Scenariot där genen nedärvt vertikalt från en gemensam förfader till Saccharomycotina och Pezizomycotina lämnar dock en del obesvarade frågor. Varför har homologer till den kloroperoxidas-liknande genen hittats både i protister, bakterier och andra organismer som inte härstammar från denna gemensamma förfader? Varför är fördelningen av knoppjästsvampar i genträdet så fläckvis? Ett svar på dessa frågor skulle kunna vara horisontell genöverföring.

Andersson (2005) beskriver hur HGT-events kan leda till en fläckvis fördelning av homologa sekvenser i ett fylogenetiskt träd. Som en förklaring till detta mönster anges något eller några enstaka HGT-event mellan bakterie och eukaryot som följs av överföring av genen eukaryoter emellan. Denna typ av fläckvisa fördelning som beskrivs av Andersson stämmer bra överens med den som genträdet i detta arbete uppvisar. Andersson menar också att något som talar för att ett HGT-event skett är när en eukaryot sekvens placerats inbäddad bland prokaryota sekvenser i ett gentråd, vilket är fallet för de knoppjästsvampar som i genträdet placerat sig utanför den större fylogenetiska gruppen, se grupp a, b, d och e samt enstaka arter av knoppjästsvampar i figur 4. Detta skulle kunna tala för att horisontell genöverföring skett någonstans i utvecklingen av dessa arter.

Det råder delade meningar om hur stor roll horisontell genöverföring haft i knoppjästsvamparnas evolution. Enligt både Shen et al. (2018) och Andersson (2005) är HGT-event ovanliga i jästsvampar, men har förekommit. Shen et al. (2018) kartlade 878 potentiella gener som blivit erhållna via sammanlagt 365 HGT-events. Majoriteten av dessa gener härstammar från bakterier. Eventen är koncentrerade framför allt till Dipodascaceae / Trichomonascaceae-kladen samt till familjen Saccharomycetaceae, men förekommer också i CUG-Ser1-kladen.

Ett oväntat resultat i det fylogenetiska trädet är fördelningen av de fyra arter inom släktet *Wickerhamiella* vars genom innehåller kloroperoxidas-likande gener. I genträdet är arterna fördelade som två åtskilda monofyletiska grupper, se grupp a och b, figur 4. De båda grupperna ligger inte direkt intill varandra, utan är separerade av noder även bakterier samt knoppjästsvampen *S. sp. 'tilneyi'* utgår ifrån.

Varför de fyra *Wickerhamiella*-arterna delar på sig och bildar två monofyletiska grupper i genträdet i stället för en skulle kunna bero på *Wickerhamiella*-arternas specialisering. Just *Wickerhamiella*-arterna har uppvisat en snäv specialisering och kan beroende på art bara livnära sig på väldigt specifika substrat, jämfört med andra släkten i Dipodascaceae/Trichomonascaceae-kladen som är mer generalistiska (Gonçalves et al. 2020). Eventuellt skulle två separata HGT-event kunna ligga till

grund för specialiseringen som gjort att dessa fyra arter bildar två monofyletiska grupper i genträdet.

Den höga frekvensen av HGT-events Shen et. al (2018) påvisade i just Dipodascaceae/Trichomonascaceae-kladen är koncentrerade kring framför allt *Wickerhamiella*-släktet. Detta pekar mot att det vore det ganska troligt att ett eller flera HGT-event är grunden till den splittrade fördelningen av *Wickerhamiella*-arterna i genträdet. Detta kräver dock vidare undersökningar för att kunna dra någon säker slutsats om.

På genträdet befinner sig de båda monofyletiska grupperna bestående av *Wickerhamiella*-arter, se grupp a och b i figur 4, långt från de grenar där de andra arterna från Dipodascaceae/Trichomonascaceae-kladen befinner sig. Gemensamt för placeringen av dessa andra arter är att de alla befinner sig i grupperingar där också bakteriesekvenser finns. Detta skulle kunna tyda på att HGT-event har påverkat utvecklingen även av dessa sekvenser, som enligt det fylogenetiska trädet har mer gemensamt med bakteriella homologer än de andra sekvenserna från Dipodascaceae/Trichomonascaceae-kladen.

De sökningar som gjordes mot knoppjästsvampar som bedömdes representativa för sina grenar i genträdet resulterade enbart i träffar bestående av bakteriella sekvenser, se tabell 4. Detta talar också för att de kloroperoxidas-liknande sekvenserna i detta projekt har sitt ursprung i ett eller flera HGT-event. Hade sekvenserna varit vertikalt nedärvt från en gemensam förfader till Saccharomycotina och Pezizomycotina hade den högsta träffen troligtvis varit en sekvens från en annan svamp, till exempel en svamp inom Pezizomycotina.

De potentiella HGT-event som enligt Shen et al. (2018) skett i CUG-Ser1-kladen är koncentrerade till två grupperingar av arter på det fylogenetiska trädet som ligger till grund för den studien. Dessa två grupper av arter korresponderar väl med de arter i CUG-ser1-kladen som uppvisar sekvenshomologerna som undersökts i detta arbete, se grupp c, figur 4. Detta bevisar ingenting på egen hand om de kloroperoxidas-liknande genernas ursprung, men visar att påverkan av HGT-events inte kan uteslutas. HGT-event må vara ovanliga hos knoppjästsvampar, men det finns ett tydligt överlapp mellan arter där horisontell genöverföring skett och där homologer till den kloroperoxidas-liknande genen påträffats.

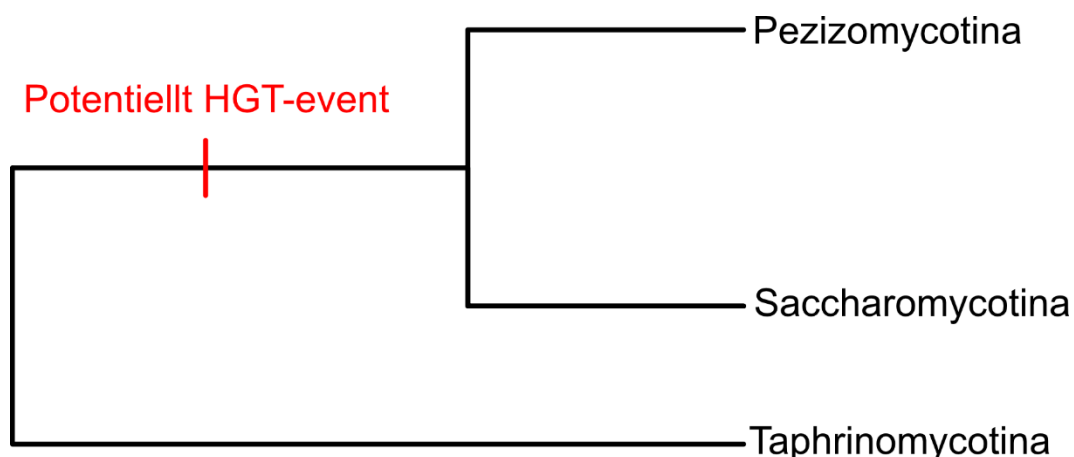
Något som bör nämnas är i samband med erhållning av gener är förekomsten av hybridisering bland knoppjästsvamparna. Hybridisering har framför allt påträffats hos knoppjästsvampar som är förknippade med mänskliga miljöer, men har också påvisats i naturen (Gabaldón 2020). Hybridisering har förekommit i både Cug-ser1-kladen och i släktet *Saccharomycopsis*. Detta skulle kunna förklara varför de båda *Saccharomycopsis*-arterna placerats sig i den stora monofyletiska gruppen c i det



fylogenetiska trädet, se figur 4, trots att de inte är taxonomiskt nära släkt med de andra arterna i den gruppen. Dessa samband är dock alldeles för svaga för att dra någon slutsats utifrån, men skulle kunna ligga till grund för ytterligare undersökningar.

Horisontell genöverföring ser ut att kunna vara en viktig aspekt i fördelningen av prokaryota sekvenser respektive sekvenser från knoppjästsvampar i genträdet. Något som inte riktigt passar in i detta mönster dock är den stora mängd homologer som påträffades för sökningar inom andra delar av riket Fungi.

Ponera att det är ett enda HGT-event som ligger till grund för erhållningen av den kloroperoxidas-liknande sekvensen till en eventuell gemensamma förfader som nämns i avsnitt 4.1. Eftersom inga homologer påträffats hos Taphrinomycotina måste detta eventuellt HGT-event alltså ha skett innan Pezizomycotina och Saccharomycotina delades till två understammar, men efter att Taphrinomycotina separerade sig från dem.



Figur 5. Schematisk bild över var ett potentiellt HGT-eventet kan skett som gjort att kloroperoxidas-liknande sekvenser erhållits av en gemensam förfader till understammarna Pezizomycotina och Saccharomycotina. Det fylogenetiska förhållandet mellan understammarna är hämtat från Cuomo & Birren (2010)

Alternativt har två mycket liknande HGT-event skett hos Pezizomycotina respektive Saccharomycotina, där överföring av en bakteriell sekvens skett till de båda understammarna oberoende av varandra, detta verkar dock orealistiskt. Troligare är att en tidig art i någon av de två understammarna erhållit sekvensen från en bakterie, och sedan i sin tur fört över den till en tidig art i den andra understammen.

Huruvida de kloroperoxidas-liknande generna blivit erhållna via flera separata HGT-event mellan bakterie och eukaryot, eller om någon enstaka överföring av genen skett mellan bakterie och eukaryot för att sedan överföras eukaryoter emellan

är inte klart. Inte heller i vilken ordning dessa event skulle skett går att säga. I och med att den stora monofyletiska i genträdet inte ligger inbäddad bland prokaryota sekvenser skulle det kunna antas att dessa arter erhållit genen genom horisontell genöverföring eukaryoter emellan eller bevarat genen från en förfader. Grupperna a,b,d och e, samt de två friliggande knoppjästsvamparna, se figur 5, ligger alla inbäddade bland prokaryota sekvenser. Detta pekar mot att ett eller flera potentiella prokaryota HGT-event skulle kunna vara källan till dessa sekvenser.

### 4.3. Scenario III – Kombination av genförlust och HGT

Både horisontell genöverföring och genförlust presenterar förlopp som kan ha skett, men ingen av mekanismerna ger var för sig svar på alla frågor. Indikationer finns på att både HGT-event genförlust skett och påverkat det evolutionära släktskapet mellan de homologa generna. Dock är det inte möjligt att dra slutsatser när och vart de olika evenen skett utifrån resultaten i detta arbete.

### 4.4. Enzymatisk funktion

Något tydligt mönster vad gäller ekologin för de arter som uppvisade homologer kunde ej upptäckas. Bland annat finns humanpatogener, växtpatogener och knoppjästsvampar förknippade med insekter bland de arter där homologer påträffats. I och med den spridda fördelningen av homologa sekvenser bland arter av knoppjästsvampar hade ett sådant mönster dock inte varit så troligt.

Inte heller går det att utifrån de arter där homologa sekvenser hittats dra någon slutsats om enzymets funktion. Troligt är att de sekvenser som kodar för enzym inom den stora monofyletiska gruppen c, se figur 4, har samma funktion då dessa sekvenser ligger tätt och få substitutioner skett.

Alla sekvenser i undersökningen uppvisar den konserverade katalytiska triad som är karaktäristisk för  $\alpha/\beta$ -hydrolaser. Detta indikerar att samtliga sekvenser kodar för funktionella  $\alpha/\beta$ -hydrolas. I samtliga fall är aminosyrorna i triaden dessutom identiska med triaden i *R. erythropolis* kloroperoxidas, vilket tyder på att sekvenserna kodar för funktionella kloroperoxidaser. För att bekräfta detta krävs dock vidare enzymologiska undersökningar.

## 4.5. Förslag på vidare forskning

Eftersom projektets fokus legat på knoppjästsvampar har resultaten från sökningar som gjorts mot andra organismtyper i hög grad berott på vilka sökningar som valdes att göras. Det finns stor möjlighet att fler peroxidas-liknande sekvenser skulle kunna upptäckas med ytterligare sökningar. Dessa undersökningar skulle kunna göras genom att fortsätta undersöka riket Fungi och framför allt divisionen Ascomycota. I och med att homologa sekvenser hittats i både Saccharomycotina och Pezizomycotina skulle det vara intressant att utforska relationen mellan den gemensamma förfadern och kloroperoxidas-liknande enzym ytterligare.

Ytterligare en aspekt som hade kunnat utökas är sökningen på de representativa sekvenser som valdes ut från det preliminära fylogenetiska trädet. Flera av de sekvenserna som sökningarna resulterades i fick gallras bort då de bedömdes otillförlitliga. Dessa sökningar hade kunnat fördjupats och tagit större del i projektet. Resultaten från dessa sökningar indikerade att en stark koppling finns mellan bakteriella kloroperoxidas-liknande sekvenser och de sekvenser som söktes med, och kanske finns fler sådana kopplingar att upptäcka.

I detta projekt hittades inga publicerade studier på funktionellt karakteriserat kloroperoxidas-likande protein hos en svamp från den genfamilj som undersöktes. Hade något sådant hittats hade det kunnat bidra till analysen. Det kan dock inte uteslutas att en sådan studie har utförts men inte påträffats i det här projektet.

För att ytterligare utöka den fylogenetiska analysen hade flera andra gener kunnat undersökas på samma sätt genom BLAST-sökningar. Även närliggande gener till den som undersöks i detta projekt hade kunnat undersökas för att eventuellt kunna ge en indikation om genens funktion.

## 4.6. Slutsats

Detta projekt har både gett svar på frågor och ställt frågor. Homologa sekvenser har hittats i knoppjästsvampar vitt åtskilda fylogenetiskt och ekologiskt samt hos en rad andra organismer som till synes inte har någon koppling med knoppjästsvampar. Kanske hade laborativa tester bekräftat att enzymet verkligen är ett kloroperoxidas, men frågorna om genens ursprung hade kvarstått.

Ett fylogenetiskt träd är i sig bara en hypotes om släktskap och ett försöka till att skapa överblick över ett förlopp som i praktiken inte är linjär tidslinje utan snarare intrikat väv av händelser. Detta projekt har visat på fylogenetikens styrkor; utifrån en enda orphan gene har en historia börjat träda fram, en historia som inte hade

synliggjorts av endast biokemiska tester av ett enzym. Trots att ingen hypotes om det undersökta enzymets funktion kunde upprättas hittades i stället tydliga indikationer på att bakterier haft en roll i enzymets utveckling. Vi vet att flertalet mekanismer så som horisontell genöverföring och genförlust förekommer och driver på utvecklingen hos knoppjästsvampar men frågorna hur, när och var kvarstår.

Att dessa enzym skulle ha samma roll hos alla vitt skilda arter av knoppjästsvampar med högst specialiserade levnadssätt är tveksamt. Något som ytterligare talar emot detta är att homologer också hittats i andra organismtyper från andra domäner.

Det går inte att dra exakta slutsatser om enzymets ursprung eller funktion i detta skede. Den enda slutsatsen som med säkerhet kan dras är att mycket finns mycket kvar att utforska och att dörren till knoppjästsvamparnas värld bara öppnats på glänt.

## Referenser

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215 (3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Andersson, J.O. (2005). Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences CMLS*, 62 (11), 1182–1197. <https://doi.org/10.1007/s00018-005-4539-z>
- Baldauf, S.L. (2003). Phylogeny for the faint of heart: a tutorial. *Trends in Genetics*, 19 (6), 345–351. [https://doi.org/10.1016/S0168-9525\(03\)00112-4](https://doi.org/10.1016/S0168-9525(03)00112-4)
- Blackwell, M. (2017). Made for Each Other: Ascomycete Yeasts and Insects. *Microbiology Spectrum*, 5 (3). <https://doi.org/10.1128/microbiolspec.FUNK-0081-2016>
- Blackwell, M. & Spatafora, J.W. (2004). 1 - FUNGI AND THEIR ALLIES. In: Mueller, G.M., Bills, G.F., & Foster, M.S. (eds.) *Biodiversity of Fungi*. Burlington: Academic Press, 7–21. <https://doi.org/10.1016/B978-012509551-8/50004-0>
- Choudhuri, S. (2014). Bioinformatics for beginners: genes, genomes, molecular evolution, databases and analytical tools. London: Elsevier.
- De Mot, R., De Schrijver, A., Schoofs, G. & Parret, A.H.A. (2003). The thiocarbamate-inducible *Rhodococcus* enzyme ThcF as a member of the family of  $\alpha/\beta$  hydrolases with haloperoxidative side activity. *FEMS Microbiology Letters*, 224 (2), 197–203. [https://doi.org/10.1016/S0378-1097\(03\)00452-X](https://doi.org/10.1016/S0378-1097(03)00452-X)
- De Schrijver, A., Nagy, I., Schoofs, G., Proost, P., Vanderleyden, J., van Pée, K.H. & De Mot, R. (1997). Thiocarbamate herbicide-inducible nonheme haloperoxidase of *Rhodococcus erythropolis* NI86/21. *Applied and Environmental Microbiology*, 63 (5), 1911–1916
- Dimitriou, P.S., Denesyuk, A., Takahashi, S., Yamashita, S., Johnson, M.S., Nakayama, T. & Denessiouk, K. (2017). Alpha/beta-hydrolases: A unique structural motif coordinates catalytic acid residue in 40 protein fold families. *Proteins: Structure, Function, and Bioinformatics*, 85 (10), 1845–1855. <https://doi.org/10.1002/prot.25338>
- Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. & Guindon, S. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59 (3), 307–321
- Gabaldón, T. (2020). Hybridization and the origin of new yeast lineages. *FEMS Yeast Research*, 20 (foaa040). <https://doi.org/10.1093/femsyr/foaa040>
- Gonçalves, P., Gonçalves, C., Brito, P.H. & Sampaio, J.P. (2020). The Wickerhamiella/Starmerella clade—A treasure trove for the study of the evolution of yeast metabolism. *Yeast*, 37 (4), 313–320. <https://doi.org/10.1002/yea.3463>
- Hanson, A.D., Pribat, A., Waller, J.C. & de Crécy-lagard, V. (2009). ‘Unknown’ proteins and ‘orphan’ enzymes: the missing half of the engineering parts list

- and how to find it. *The Biochemical journal*, 425 (1), 1–11. <https://doi.org/10.1042/BJ20091328>
- Hittinger, C.T., Rokas, A., Bai, F.-Y., Boekhout, T., Gonçalves, P., Jeffries, T.W., Kominek, J., Lachance, M.-A., Libkind, D., Rosa, C.A., Sampaio, J.P. & Kurtzman, C.P. (2015). Genomics and the making of yeast biodiversity. *Current Opinion in Genetics & Development*, 35, 100–109. <https://doi.org/10.1016/j.gde.2015.10.008>
- Hofmann, B., Tölzer, S., Pelletier, I., Altenbuchner, J., van Pée, K.H. & Hecht, H.J. (1998). Structural investigation of the cofactor-free chloroperoxidases. Edited by R. Huber. *Journal of Molecular Biology*, 279 (4), 889–900. <https://doi.org/10.1006/jmbi.1998.1802>
- Hofmann, K. & Baron, M. (070521). *BoxShade*. Version: 3.21. [https://embnet.vital-it.ch/software/BOX\\_form.html](https://embnet.vital-it.ch/software/BOX_form.html)
- Huster, M., Müller-Renno, C., Ziegler, C., Schlegel, C., Ulber, R. & Muffler, K. (2016). Chloroperoxidase production by *Caldariomyces fumago* biofilms. *Engineering in Life Sciences*, 16 (1), 88–92. <https://doi.org/10.1002/elsc.201500032>
- Krassowski, T., Coughlan, A.Y., Shen, X.-X., Zhou, X., Kominek, J., Opulente, D.A., Riley, R., Grigoriev, I.V., Maheshwari, N., Shields, D.C., Kurtzman, C.P., Hittinger, C.T., Rokas, A. & Wolfe, K.H. (2018). Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. *Nature Communications*, 9 (1), 1887. <https://doi.org/10.1038/s41467-018-04374-7>
- Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35 (6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Kurtzman, C. (2011). *The Yeasts A Taxonomic Study*. Burlington: Elsevier Science.
- Lefort, V., Longueville, J.-E. & Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution*, 34 (9), 2422–2424. <https://doi.org/10.1093/molbev/msx149>
- Mindrebo, J.T., Nartey, C.M., Seto, Y., Burkart, M.D. & Noel, J.P. (2016). Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Current Opinion in Structural Biology*, 41, 233–246. <https://doi.org/10.1016/j.sbi.2016.08.005>
- Proux-Wéra, E., Armisen, D., Byrne, K.P. & Wolfe, K.H. (2012). A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics*, 13 (1), 237. <https://doi.org/10.1186/1471-2105-13-237>
- Shen, X.-X., Opulente, D.A., Kominek, J., Zhou, X., Steenwyk, J.L., Buh, K.V., Haase, M.A.B., Wisecaver, J.H., Wang, M., Doering, D.T., Boudouris, J.T., Schneider, R.M., Langdon, Q.K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R., Čadež, N., Libkind, D., Rosa, C.A., DeVirgilio, J., Hulfachor, A.B., Groenewald, M., Kurtzman, C.P., Hittinger, C.T. & Rokas, A. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, 175 (6), 1533–1545.e20. <https://doi.org/10.1016/j.cell.2018.10.023>
- Uniprot *thcF* - Non-heme haloperoxidase - *Rhodococcus erythropolis* - *thcF* gene & protein. <https://www.uniprot.org/uniprot/O05691> [2021-06-09]

# Tack

Stort tack till min handledare Tomas som guidat mig genom det här arbetet i alla zoom-timmar. Jag kan på riktigt inte tänka mig ett roligare kandidatprojekt. Du har lärt mig allt jag idag kan om bioinformatik, och lite till som jag fortfarande inte förstår.

Tack till Simon för virtuellt kontor och för crash course i Affinity designer, och tack till Johan för korrläsning och kaffekokning.

## Bilaga 1

Flersekvensjämförelse av kloroperoxidas-liknande proteinsekvenser som användes för att konstruera det fylogenetiska trädet i denna studie. Knoppjästsvamparterna *Debaryomyces fabryi*, *Debaryomyces subglobosus*, *Debaryomyces prosopidis*, *Meyerozyma* sp. JA9, *Meyerozyma caribbica*, *Meyerozyma athensensis*, *Priceomyces medius*, *Candida duobushaemulonis*, *Candida pseudohaemulonii*, *Blastobotrys adenivorans* och *Wickerhamiella infanticola* är exkluderade ur figuren då släkten är representerade av andra arter med snarlika sekvenser. Rosa markering indikerar identisk position, turkos markering indikerar liknande position, minst 70%. Positionerna för de aminosyror som ingår i den katalytiska triaden serin-asparagin-histidin är markerade med gult.

|        |             |
|--------|-------------|
| Debhan | -----M----- |
| Meygui | -----M----- |
| Cancar | -----M----- |
| Cantam | -----M----- |
| Pricar | -----M----- |
| Canasc | -----M----- |
| Wicano | -----M----- |
| Canint | -----M----- |
| Canbla | -----M----- |
| Canaur | -----M----- |
| Canhae | -----M----- |
| Sacfod | -----M----- |
| Hypbur | -----M----- |
| Debmar | -----M----- |
| Hyphei | -----M----- |
| Cangot | -----M----- |
| Serodo | -----M----- |
| Ser_sp | -----M----- |
| Hyphom | -----M----- |
| Cangor | -----M----- |
| Sacmal | -----M----- |
| Canaas | -----M----- |
| Ogapol | -----M----- |
| Tricif | -----M----- |
| Blamok | -----M----- |
| Sugxyl | -----M----- |
| Blaser | -----M----- |
| Exooli | -----M----- |
| Blaraf | -----M----- |
| Phicur | -----M----- |
| Speeur | -----M----- |
| Molsco | -----M----- |
| Hel_sp | -----M----- |
| Scylig | -----M----- |
| Cadmal | -----M----- |
| Cad_sp | -----M----- |
| Claban | -----M----- |
| Clapsa | -----M----- |
| Fonnub | -----M----- |
| Fonped | -----M----- |



|         |  |
|---------|--|
| Fonmon  | -----M-----                                      |
| Claimm  | -----M-----                                      |
| Fonere  | -----M-----                                      |
| Rhilig  | -----M-----                                      |
| Mycgal  | -----M-----                                      |
| Mycven  | -----M-----                                      |
| Neodit  | -----M-----                                      |
| Exoaqu  | -----M-----                                      |
| Fonmul  | -----M-----                                      |
| Cypeur  | -----M-----                                      |
| Polcit  | -----M-----                                      |
| Phiatt  | -----M-----                                      |
| Lacocc  | -----MSLTVPSLFNSTPFVPPI-----                     |
| Pseflu  | -----  |
| Syn_sp  | -----M-----                                      |
| Metace  | -----M-----                                      |
| Metsil  | -----M-----                                      |
| Methor  | -----M-----                                      |
| Met_sp  | -----M-----                                      |
| Pedpar  | -----M-----                                      |
| Metthe  | -----M-----                                      |
| Wicsor  | -----M-----                                      |
| Novros  | -----M-----                                      |
| Wicdom  | -----M-----                                      |
| Rhoery  | -----M-----                                      |
| Burpyr  | -----M-----                                      |
| Wicocc  | -----M-----                                      |
| Strliv  | -----  |
| Starrat | -----M-----                                      |
| Burcen  | -----GPGSM-----                                  |
| Pseput  | -----M-----                                      |
| Sta_sp  | MHVCSQRVLPALILLAAPDISTARTVVTSLAQSPLLRNVQLEM----- |

|         |  |
|---------|--|
| Debhan  | -----TGSAKLKLSDGTIFYKDWGT--G-DPIVFSHGWPLTSDAFEDQMMFLALKGFR     |
| Meygui  | -----TTSSKLTLSGDGTIFYKDWGT--G-EPIVFSHGWPLTSDAFEDQMIYLALKGYR    |
| Cancar  | -----PRLTSLSDGTIFYKSWGT--G-DPIVFSHGWPLTSDAFEDQMLYLASKGYR       |
| Cantam  | -----SSKLILKDGTEIFYKDWGK--G-QPIVFSHGWPLTSDAFEDQMYYLALKGFR      |
| Pricar  | -----TGSaelVLSDGTLIFYKDWGS--G-EPIVFSHGWPLTSDAFEDQMMSLASKGFR    |
| Canasc  | -----PTLRLSDGTIFYKDWGT--G-QPVVFSHGWPLSSDAFEDQMLYLASKGFR        |
| Wicano  | -----PYLKLSDGTDLYYKDWGT--G-DPIVFSHGWPLNADAFEDQMIYLANKGFR       |
| Canint  | -----PYLKLSDGTDLYYKDWGT--G-TPVVFSSHGWPLSADAFEDQMIFLALKGYR      |
| Canbla  | -----PYLKLSDGTDLYYKDWGT--G-DPIVFSHGWPLSADAFEDQMFLYLASKGYR      |
| Canaur  | -----SGSSYLKLSDGTNLIFYKDWGT--G-DPIVFSHGWPLSSDAFEDQMFLYLASKGYR  |
| Canhae  | -----GGSSYLKLSDGTDIFYKDWGT--G-DPIVFSHGWPLSSDAFEDQMFLYLASKGYR   |
| Sacfod  | -----PYFETSDGTSLIFYKDWGH--G-DAVVFSHGWPLSSDAFEDQMLFLANKGYR      |
| Hypbur  | -----GSSILKLSDGTDLYYKDWGV--G-EPVVFSSHGWPLSSDAFEDQMIYLALKGYR    |
| Debmar  | -----NSATLKLSDGTIFYKDWGN--G-DPIVFSHGWPLTSDAFEDQMIFLASKGYR      |
| Hyphei  | -----SGSAKLTLLDGTIFYKDWGV--G-EPIVFSHGWPLSSDAFEDQMIFLASKGYR     |
| Cangot  | -----SGSATLKLSDGTIFYKDWGV--G-EPIVFSHGWPLSSDAFEDQMIFLASKGYR     |
| Serodo  | -----SNSTTLKLKDGTDLYYKDWGE--G-QPVVFSHGWPLSGDAFEDQMLFLASHGYR    |
| Ser_sp  | -----NNSNFLKLKDGTHLIFYKDWGS--G-QPVVFSHGWPLTSDAFEDQMLFLANNGYR   |
| Hyphom  | -----TGSSTLTLSGDGTIFYKDWGY--G-GPIVFSHGWPLSSDAFEDQMYYLASRGYR    |
| Cangor  | -----SNYLTLDGAEIFYKDWGNPKG--TPIVFSHGWPLSSDAFEDQMLFLANKGYR      |
| Sacmal  | -----SGSNTLKLFDGTIFYKDWGK--G-DPIVFSHGWPLTSDAFEDQMYYLANKGFR     |
| Canaas  | -----QNSSFLSLKDGTLFFKDWGT--G-QPIVFHGWPLSSDAFEDQMLFLANKGYR      |
| Ogapol  | -----PFLKLSDGAELIFYKDWGNADG--EIVTFSHGWPLSSDNWENQMFLANKGYR      |
| Tricif  | -----PVIKLRLDGAELIFYKDWGNPKG--PIVTFSHGWPLSSDNWENQMFLASHGYR     |
| Blamok  | -----PYLKLRLDGAELIFYKDWGNPKG--EIVTFSHGWPLSSDNWENQMFLADHGYR     |
| Sugxyl  | -----PYLKLRLDGAELIFYKDWGNPDG--PIVTFSHGWPLSSDNWENQMFLADHGYR     |
| Blaser  | -----PYLKLRLDGAELIFYKDWGNPKG--PVVTFSHGWPLSSDNWENQMIFLGHKGYR    |
| Exooli  | -----PHLKLRLDGAELIFYKDWGNPKG--PIVTFSHGWPLSSDNWENQMIFLASHGYR    |
| Blaraf  | -----PYLKLRLDGAELIFYKDWGNPKG--EIVTFSHGWPLSSDNWDSQLLYLGDKGYR    |
| Phicur  | -----PLLRLRLDGAELIFYKDWGNPKG--AIVTFSHGWPLSSDNWENQMAFLGEGGYR    |
| Speeur  | -----PFITTKDQTEIYYKDWGNPKG--QPVVFSHGWPLSSDNWENQMFFLANHGYR      |
| Molsco  | -----PFVTTSDKTEIYYKDWGNPKG--KPVVFSHGWPLSSDNWENQMFFLASHGYR      |
| Hel_sp  | -----PFVTTSDKTEIYYKDWGNPKG--KTVVFSHGWPLSSDNWENQMFFLSSQGYR      |
| Scyliy  | -----PYITTRDKTEIYYKDWGNPKG--QPVVFSHGWPLSSDNWESQMFFLSNQGYYR     |
| Cadmal  | -----PFITTKDQTSIFYKDWGNPKG--QVVTFSHGWPLSSDNWESQMFFLASKGYR      |
| Cad_sp  | -----PFITTKDQTSIFYKDWGNPKG--QVVTFSHGWPLSSDNWESQMFFLASKGYR      |
| Claban  | -----PYLTTNDGTEIYYKDWGHKADAPATVVFSHGWPLSSDNWENQMFFLASKGYR      |
| Clapsa  | -----PYLTTNDGTEIYYKDWGHKADAPATVVFSHGWPLSSDNWENQMFFLASKGYR      |
| Fonnu   | -----PYLTTNDGTEIYYKDWGHKADASGTVVFSHGWPLSSDNWESQMFFLASKGYR      |
| Fonped  | -----PYLTTNDGTEIYYKDWGHKADASGTVVFSHGWPLSSDNWESQMFFLASKGYR      |
| Fonmon  | -----PYLTTNDGTEIYYKDWGHKADASGTVVFSHGWPLSSDNWESQMFFLASKGHR      |
| Claimm  | -----PYLTTNDGTEIYYKDWGHKADASGTVVFSHGWPLSSDNWENQMFFLASKGYR      |
| Fonere  | -----PYLTTNDGTEIYYKDWGNKADASATVVFSHGWPLSSDNWESQMFFLASKGYR      |
| Rhilig  | -----PYITTKDGTIYYKDWGNKNG--AVVTFSHGWPLSSDNWENQMFFLASKGYR       |
| Mycgal  | -----PYLTLKDGVEIYYKDWGNKNG--PTVVFSHGWPLSSDNWENQMFFLANQGCR      |
| Mycven  | -----PYLTLKDGVEIYYKDWGNKNG--PTVVFSHGWPLSSDNWENQMFFLANQGCR      |
| Neodit  | -----PTFTLEDGVEIYYKDWGNPKG--KTVTFSHGWPLSSDNWDSQMLFLASQGYR      |
| Exoaqu  | -----PYITTHDGTEIYYKDWGSKTA--QPVVFSHGWPLSSDDWEHQMFFLASKGYR      |
| Fonmul  | -----PYITTDGTEIYYKDWGNKSA--KPITFSHGWPLSSDDWENQMFLANKGYR        |
| Cypeur  | -----PYITVKDGTEIYYKDWGNKSG--QPVVFSHGWPLSSDDWENQMFFLANKGYR      |
| Polcit  | -----PYVTAQDGTQIYYKDWGNKGS--QPVVFSHGWPLSSDDWENQMFFLAENGYYR     |
| Phiaatt | -----PFFVTQDGTIYYKDWGNKSA--PVVTFSHGWPLSSDDWDSQATFLVNKGYR       |
| Lacocc  | STPQFNNIGSYITTPDNTQIYYKDWGNPCG--EPIVFSHGWPLSSDNWENQMFFLSNEGYYR |
| Pseflu  | -----TTFTTRDGTQIYYKDWGS--G-QPIVFSHGWPLNADSWESQMIFLAAQGYR       |
| Syn_sp  | -----STITTKDGTIYYKDWGS--G-QPIVFSHGWPLSADAWESQMFFLASHGYR        |
| Metace  | -----TAITTKDGTIYYKDWGS--G-QPVVFSHGWPLSSDAFEDQMFFLASKGYR        |
| Metsil  | -----TAITTKDGTIYYKDWGS--G-QPVVFSHGWPLSSDAFEDQMFFLASKGYR        |
| Methor  | -----TTITTEDGTQIYYKDWGS--G-QPVVFSHGWPLSADAFEDQMFFLASKGYR       |
| Met_sp  | -----NMITTKDGTQIYYKDWGS--G-QPVVFSHGWPLSADAFEDQMFFLASKGYR       |
| Pedpar  | -----NTITTKDGTQIYYKDWGT--G-QPIVFSHGWPLSADAFEDQMFFLASKGYR       |
| Metthe  | -----PTIRMKDNTQIYYKDWGS--G-QPVVFSHGWPLSSDAWEDQMFFLAHGYR        |
| Wicsor  | -----PSITTSKTEIFYKDWGSKDA--HPIMFHHGWPLSSDDWDQMLFFVNQGYR        |
| Novros  | -----PFITTTDGTIFYKDWGSRDA--QPIVFHHGWPLSSDDWDAQMMFFLAKGYR       |
| Wicdom  | -----TFITTTQDGTIFYKDWGSKDA--QPIVFHHGWPLSSDDWDAQMLFFLQDQFR      |
| Rhoery  | -----PFVTASDGTIFYKDWGS--G-RPIMFHHGWPLSSDDWDSQLLFLVQRGYR        |
| Burpyr  | -----PYVTTKDNVEIFYKDWGPKDA--QPIVFHHGWPLSSDDWDAQMLFFVQKGYR      |
| Wicocc  | -----GFVKTDDVDIFYKDWGSKDA--QPIVFHHGWPLSSDDWDNQMLFYQKGYR        |
| Strliv  | -----GTVTTSDDGTIFYKDWGPRDG--LPVVFHHGWPLSSDDWDAQMLFFLASHGYR     |
| Starrat | -----GFVTKDNVEIFYKDWGSKDA--PVIYFHHGWPLSSDDWDAQMLFFVNNGFR       |
| Burcen  | -----GTVTTKDGTIFYKDWGPRDA--KVIHFHHGWPLSSDDWDAQMLFFVNAHGYR      |
| Pseput  | -----SYVTTKDGTIFYKDWGPRDA--PVIHFHHGWPLSSDDWDAQMLFFLAHGYR       |
| Sta_sp  | -----PFVETADGVEIYYKDWGSSES--PVIFFHHGWPLSSDDWDQMLFFLHKGFR       |

|         |  |
|---------|--|
| Debhan  | VIAHRRGHGRSSQPWEGHNTMDQYADDLAEIVDHLGLKNAIHVGHSTGGGEVARYIGRHG |
| Meygui  | VVAHRRGHGRSSQPWNHNMNDQYADDLAEIVKHLDLKNAVHVGHSTGGGEVARYISRHG  |
| Cancar  | VVAHRRGHGRSSQPWNHNTMDQYADDLAEIVKHLDLKNAVHVGHSTGGGEVARYIARHG  |
| Cantam  | VIAHRRGHGRSSQPWDGHNMDQYADDLAEIVNHLDLKNAVHGHSTGGGEVARYIARHG   |
| Pricar  | VIAHRRGHGRSSQSWEDHTMDQYADDLAEIVHLKLVKAVHGHSTGGGEVARYIGRHG    |
| Canasc  | VIAHRRGHGRSSQPWEGHNMDQYADDLSELVNLKLDKAVHGHSTGGGEVARYIGRHG    |
| Wicano  | VIAHRRGHGRSSQPWDGNNLDQYADDLAEIVHHLKLVNVHVGHSTGGGEVARYIGRHG   |
| Canint  | VIAHRRGHGRSSQPWDGHNMDQYADDLAEIVKHLDLKNAIHVGHSTGGGEVARYVGRHG  |
| Canbla  | VIAHRRGHGRSSQPWDGHNMDQYADDLAEIVKHLGLKNAVHVGHSTGGGEVAKFIARHG  |
| Canaur  | VIAHRRGHGRSSQPWSGHNMDQYADDLAEIVNHLGLKNAVHVGHSTGGGEVARYLARHG  |
| Canhae  | VIAHRRGHGRSSQPWDGHNMDQYADDLAEIVKHLDLKDAVHVGHSTGGGEVARYLARHG  |
| Sacfod  | VIAHRRGHGRSSQPWDGHNMDQYADDLAEIVKHLGITNAVHVGHSTGGGEVARYIGRHG  |
| Hypbur  | VIAHRRGHGRSSQSWDGHNMDQYADDLAEIVNHLDSLKAVHGHSTGGGEVARYVARHG   |
| Debmar  | VIAHRRGHGRSSQPWEGHNMDQYADDLAEIVNHLDLKTAIHGHSTGGGEVARYIGRHG   |
| Hyphei  | VIAHRRGHGRSSQPWNHNMNDQYADDLAEIVKHLDLKDAVHGHSTGGGEVARYIGRHG   |
| Cangot  | VIAHRRGHGRSSQQTWDGHNMDQYADDLAEIVKHLDLKNVIHGHSTGGGEVARYIGRHG  |
| Serodo  | VIAHRRGHGRSSQPWDGHNMDQYADDLAEITAAIDLQNAVHVGHSTGGGEVARYIARHG  |
| Ser_sp  | VIAHRRGHGRSSQPWDGHNMDQYADDLAEITAHLDLKDVAHVGHSTGGGEVARYIGRHG  |
| Hyphom  | VIAHRRGHGRSSQPWTGHTMDQYADDLAEIVNHLHLKAVLVGHSTGGGEVTRYVARHG   |
| Cangor  | VIAHRRGHGRSSQPWEGHNMDQYADDLAEIVNHLNLKDAIHGHSTGGGEVARYIGRHG   |
| Sacmal  | VIAHRRGHGRSSQSFSGHTLDQYADDLAEIMKHLNITNAVHVGHSTGGGEVARYIARYG  |
| Canaas  | VIAHRRGHGRSSQAFNGHNMDQYADDVAELVAHLNLKNVIHGHSTGGGEVARYIARQG   |
| Ogapol  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Tricif  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Blamok  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Sugxyl  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Blaser  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Exooli  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Blaraf  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Phicur  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Speeur  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Molsco  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Hel_sp  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Scyliy  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Cadmal  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Cad_sp  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Claban  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Clapsa  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Fonnub  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Fonped  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Fonmon  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Claimm  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Fonere  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Chilig  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Mycgal  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Mycven  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Neodit  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Exoaqu  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Fonmul  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Cypeur  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Polcit  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Phiaatt | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Iacocc  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Pseflu  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Syn_sp  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Metace  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Metsil  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Methor  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Met_sp  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Pedpar  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Metthe  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Wicsor  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Novros  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Widom   | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Rhoery  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Burpyr  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Wicocc  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Strliv  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Starat  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Burcen  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Pseput  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |
| Sta_sp  | VIAHRRGHGRSSQTFWDDGNNMTFVDDLEELFKHLQVDAVMVGHSHGGGEVTHYLGKHG  |

Debhan  
 Meygui  
 Cancar  
 Cantam  
 Pricar  
 Canasc  
 Wicano  
 Canint  
 Canbla  
 Canaur  
 Canhae  
 Sacfod  
 Hypbur  
 Debmar  
 Hyphei  
 Cangot  
 Serodo  
 Ser\_sp  
 Hyphom  
 Cangor  
 Sacmal  
 Canaas  
 Ogapol  
 Tricif  
 Blamok  
 Sugxyl  
 Blaser  
 Exooli  
 Blaraf  
 Phicur  
 Speeur  
 Molsco  
 Hel\_sp  
 Scylig  
 Cadmal  
 Cad\_sp  
 Claban  
 Clapsa  
 Fonhub  
 Fonped  
 Fonmon  
 Claimm  
 Fonere  
 Rhilig  
 Mycgal  
 Mycven  
 Neodit  
 Exoaqu  
 Fonmul  
 Cypeur  
 Polcit  
 Phiatt  
 Lacocc  
 Pseflu  
 Syn\_sp  
 Metace  
 Metsil  
 Methor  
 Met\_sp  
 Pedpar  
 Metthe  
 Wicsor  
 Novros  
 Wicdom  
 Rhoery  
 Burpyr  
 Wicocc  
 Strliv  
 Starrat  
 Burcen  
 Pseput  
 Sta\_sp

I--GKVS KCVLISAVPPLMVKSEENLVGTPIEVFHDIRDNVSKDRSKFFKDL-AFTFYGL  
 T--LRVLKAVLLGAIPPLMLKTDANPSGTPKEVFDGIRANVLKDRSKFFKEL-ALSFFYGF  
 T--SRVSKAVLLGAIPPLMLKTAANPSGTPKQVFDGIRENVKDRSKFFKEL-ALGFFYGF  
 T--SRVSKAVLLGAIPPLMVKTASNPDGTPVEVFDGIRDNVKDRSTFFKEL-ALTFFYGF  
 T--SRVSKAILIGAIPPLMIKTENPLGTPVEVFDGIRSNVLNDRSEFFKEL-ASSFFYGF  
 T--SKVSKAVLLGAIPPLMVKSESNPEGTPVEVFDGIRSNVLKDRSQFFKEL-AKSFFYGF  
 V--NNVSKAVLLGAIPPLMLKTESNPNGTPIEVFDGIRSNVLKDRSKFFKEL-APTFFYGF  
 T--DRVQKAVLLGAIPPLMLKTSNPNGTPIVFDGIRSNVLKDRSQFFKDL-APAFYGF  
 N--ENVSKAILIGAIPPLMLKTASNPDGTPIDVFDGIRSNLLKDRSSFFKEL-ALAFYGF  
 T--KNARKAVLLGAIPPLMIKTDFNPDGVPKDALDDIRNVKDRSQFFQDL-PTLFYGF  
 T--SRASKAVLLGAIPPLLIQTDFNPQGVPREALDDIRNVKDRSQFFKDL-PNLFYGF  
 T--KNVKAVLLGAIPPLMLKTESNPNGTPIVFDGIREGVVDRSQFFKDL-AITFYGF  
 T--QRLKAILIGAIPPLMLKSDNANGTPIEVFDEIRNNVLQDRSQFFQDL-APVFFYGF  
 A--EKVSKAVLLGAIPPLMVKSESNPEGTPIGVFDGIRSNVLKDRSQFFKDL-APVFFYGF  
 T--ERVSKAILIGAIPPLMAKTEFNPGGLPIENFDGIRNGVLTDRSQFFKDL-APVFFYGF  
 T--SRVSKAILIGAIPPLMLKSEINPQGTPIIDTFDNIRKVLDRSQFFKEL-ASAFYGF  
 T--ERVAKAVLLGAIPPLMLKTDNFNPGGLPLSVFDGIRGVDVNDRSQFFKDL-SETFYGF  
 T--SRVAKAVLLGAIPPLMIKTDFNPDGLPLSVFDGIRGVDKNDRSQFFRDL-TQAFYGF  
 T--ERISKVLLGAIPPLMLKTDANPKGTPIVFDGIRASVLKDRSAFFKQL-AVGFFYGL  
 T--SRVQKAVLLGAIPPLMIKTSNPNGTPIVFDGIRSNVLKDRSGFFKEL-ALVFFYGF  
 T--KDVKCVLLGAIPPLMLKTDGNPDGPKDVFDRIQNVLDNRSSFFKEL-AMFFYGF  
 T--SNVKAAILIGAIPPLMLKTEQNPNGLPPIEVNDFRNVLTNRAEFFKEA-AIPFFYGF  
 T--GRVKKAVLLCAVPPPLMVKTASNPDGLDRSVFDGIFREAMHKDRAQFFLDIPRGFFYGF  
 T--SRVKKAVLLCAVPPPLMVKTSSNPEGLDKSVFDSFREAMHKDRAQFFLDVPSGFFYGF  
 T--SRVKKAVLLCAVPPPLMVKTNNPEGTDISVFDGIFREAMHKDRAQFFLDVPSGFFYGF  
 T--SRVKKAVLLCAVPPPLLVKTEANKEGIDVSFDSFRKQILEDRAQFFLDVPSGFFYGF  
 T--SRVKKAVLLCAVPPPLLLKTPENPEGLDKSVFDSFREAMKKDRSQFFLDVPTGFFYGF  
 T--ARFKKAVLLCAVPPPLMVKTSSNPEGTDKSVFDSFREAMLQDRAQFFLDVPSGFFYGF  
 T--GRFKKAVLLCAVPPPLLLKTSNPGEVDKSVFDTFRKSLLEDRAQFFLDIPSGFFYGF  
 T--GRVRKAVLLCAVPPPLMLKTAANPEGTDKAVFDSFRDAMRRDRAQFFLDVATGFFYGF  
 T--SRVSKAVLLSSVPPVMVKKASNPDGLPIDVFDGFRKAMEKDRAQFFLDVPTGFFYGF  
 T--SRVSKAVLLCAVPPPLMVKTSSNPKGLPIVFDGFRSAMIQDRAQFFLDVPSGFFYGF  
 T--SRVSKAVLLSAVPPPLMVKTEANPKGLPIEVFDGFRSAMIQDRSQFFLDVPTGFFYGF  
 S--SRVAKAVLLSAVPPVMVKKESNPNGTPIEVFDSFRAMEKDRAQFFLDVPTGFFYGF  
 S--SRVSKAVLLSAVPPVMVKKESNPNGTPIEVFDSFRAMIADRAQFFLDVPSGFFYGF  
 S--SRVSKAVLLSAVPPVMVKKESNPNGTPIEVFDSFRSAMIADRAQFFLDVPSGFFYGF  
 S--SRVKKAVLLSAVTPLMLRTEANPKGTPIEVFDSFRVAMIRDRSQFFLDLPSGFFYGF  
 S--SRVKKAVLLSAVTPLMLKTEANPKGTPIEVFDSFRAMIADRAQFFLDVPSGFFYGF  
 S--SRVKKAVLLSAVTPLMLKTEANPKGTPIEVFDSFRAMIADRAQFFLDVPSGFFYGF  
 S--SRVKKAVLLSAVTPLMLKTEANPKGTPIEVFDSFRAMIADRAQFFLDVPSGFFYGF  
 S--SRVKKAVLLSAVTPLMLKTEANPKGTPIEVFDSFRAMIADRAQFFLDVPSGFFYGF  
 S--SRVKKAVLLSAVTPLMLKTEANPKGTPIEVFDSFRAMIADRAQFFLDVPSGFFYGF  
 S--SRVKKAVLLSAVTPLMLKTEANPKGTPIEVFDSFRAMIADRAQFFLDVPSGFFYGF  
 A--SRVSKAVLLSAVTPLMLKTEANPKGTPIEVFDSFRAMIADRAQFFLDVPSGFFYGF  
 T--SRVSKAVLLSSVPPMLKTDKNPGGLSIEVNSFRDATTKDRSQFFLDVPSGFFYGF  
 T--SRVSKAVLLSAVPPVMVQSESNPEGTPLAVIDGFRAMEKDRAQFFLDVPSGFFYGF  
 T--SRVSKAVLLSAVPPVMVQSESNPEGTPLAVIDGFRAMEKDRAQFFLDVPSGFFYGF  
 T--SRVSKAVLLSAITPLMVQTAANPEGLPLGVFDSFREAMLKDRSQFFLDVPSGFFYGF  
 T--SRVSKAVLLSAVAPPLMLKTEANPDGVPLEVFDGFRDAMVDRSQFFLDVPSGFFYGF  
 T--SRVSKAVLLSAVTPGILKTDANPDGVPLEVFDGFRASMLKDRSQFFLDVPSGFFYGF  
 T--KRVSKAVLLSATAPGLLKSNDPHGVMDVFSFRDAMVDRSQFFLDVPSGFFYGF  
 S--KRVSKAVLLGAVPPLLVKSEKSPDGPVLEVFDGFRDAMVDRSQFFLDVPSGFFYGF  
 T--SRVSKAVLLSAVAPGLKTDNATGVDISVFDGFRAMNDRSQFFLDVPSGFFYGF  
 T--SRVSKAVLLSSVPPMLQSESNPVGVPMVSFDRSMTKNRAQFFLDVPTGFFYGF  
 T--ARVAKAGLLSAVPPPLMLKTEANPGGLPMEVFDGIRQASLADRSQYKDLASGFFYGF  
 T--QRVSKAVLLMGAIPPLMLKTEANPGGLPIEVFDGFRAAFLADRSQFFLDVASGFFYGF  
 T--DRVAKAVLLGAIPPLMLKTPANPAGTPIEVFDQIRASVADRQYWKDL-SLPFFYGY  
 T--DRVARAVLLGAVSPLMLKTPANPAGTPIEVFDQIRASVADRQYWKDL-SLPFFYGY  
 T--KRVAKTVLLGAIPPLMLKTEANPAGTPIEVFDQIRASVADRQYWKDL-SLPFFYGY  
 T--KRVAKAVLLGAIPPLMLKTAANPGGTPIEAFDQIRAAVLADRSQYWKDL-SMPFFYGY  
 T--KRVAKAVLLGAIPPLMLKTAANPAGTPIEAFLQIRAAVLADRSQYWKDL-SMPFFYGY  
 T--DRVAKVLLIDAVPPFLKTAANPGGIPRETDFEIRAAVLADRSQYWKDL-SMPFFYGY  
 P--GRVSKAVLLSAVTPEMVQSDINPAGTPKKDLNIREQVATNRSKFFYDF-TLPFFYGF  
 P--GRVAKAVLLSAVSPIMVQSESNPEGVPMVSFDMVREQTATNRSKFFYDF-TLPFFYGF  
 EPQGRVAKAVLLSSVPPPLMVKTENNNTDGTPEVFDQFRAGLAANRSQYLDVASGFFYGY  
 A--GRVAKAVLLGAIPPLMVQTESNPEGLPVEVFDGFRAAVVTNRSQYLDLASGFFYGY  
 QPAGRVAKAVLLSAVPPPLMLKTESNPEGLPIDVFDGFRKALADNRAQFFLDVPSGFFYGF  
 EPQGRVAKAVLLSSVPPPLMVQTKENPDGVPISVFDGFRASLAADRAAFYLDVPSGFFYGF  
 P--GRVAKAVLLSAVPPPLMVKTSSNPNGTPIEVFDGFRAAVVTNRSQYLDVPSGFFYGY  
 E--SKVSKAVLLSSVPPPLMVKSESNPEGTPKSVFDEFREALAANRSQYLDVPSGFFYGY  
 E--RNVSKAVLLSSVPPPLMVKTSSNPNGTPIEVFDGFRAAVVTNRSQYLDVPSGFFYGY  
 E--DKVAKAVLLIAAVPPPLMVQTPGNPGGLPKSVFDFGQAQVANSRAQYLDVPSGFFYGY  
 E--HDVKKAVLLSSVPPPLMVKTANNPNMGMEKSIFFDQFQVATNRPQYLDVPSGFFYGY



[illegible]

|         |   |
|---------|---|
| Debhan  | VPPISSSKAAIKFLPQAKEKVYAGAPHGLCSTHKKLVNEDIYRFITH-----          |
| Meygui  | VPPGPSSEAAIKLLPQATEKVYPGASHGLCSTHKDQVSKDIYEFIKQ-----          |
| Cancar  | VPPGPSKAAVKLLPQATEKVYPGASHGFCSTHKDEVSKDIYEFITK-----           |
| Cantam  | VPPGASSKAAVKLLPKSVKVPYPGAPHGLCSTHKDQVNDLYEFIKS-----           |
| Pricar  | VPPAASSEEAIKLLPQATKKVYVGASHGLCSTHKKDINKDIYDFITR-----          |
| Canasc  | VPPAASSKEAIKLLPSATEKVYPGAPHGLCSTHKQDVNKDIYDFIREK-----         |
| Wicano  | VPPIASSQAALKLLPQPTKVKYKGASHGVCSTHKDQVSEDILKFIKN-----          |
| Canint  | VPPVASSSEAAVKLLPQGTSEKVPYKGASHGLCTTHKDQINEDLLAFIEK-----       |
| Canbla  | VPPVASSQAALKLLSKGSEKVPYKGASHGLCSTHKDEINKDIYDFISK-----         |
| Canaur  | VPPKTSSEAAIKLLPKGEKKVYEGGSHGICSTHKDQVNDLYEFISK-----           |
| Canhae  | VPPVTSSQAALKLLPKGSEKVPYKGSGHICSTHKDEVSKDLYDFITK-----          |
| Sacfod  | VPPVSSSEAAIKLLPNGEKGVYTGSGHGLCTIDKERINEDLLAFIKK-----          |
| Hypbur  | VPPIASSQAALKLLPLPTEKVYKGASHGVCSTHKDEVNQDILAFIQ-----           |
| Debmar  | VLPVSSQAALKLLPLPTEKVYIGASHGLCSTHKDAVNEDILSFITH-----           |
| Hyphei  | VPPISSQAALKLLPQPTKVIYKGASHGVCSTHKNQVNEDILAFIKQ-----           |
| Cangot  | VPISSSRAALKLLPIPTFIVYKGGSHGICSTHKDQINKDILAFIQQ-----           |
| Serodo  | VPPISSSQAALKLLPNAQEKVYPGASHGLCSTHKAQINQDILLDFIQR-----         |
| Ser_sp  | VPPVSSSKAALKLLPQGTSEKVYPGASHGLCSTHKDEINQDILLAFIRS-----        |
| Hyphom  | VPPIASSEAAIKLLPHGTEKVYPGASHGVCSTHKDEVSKDIYDFIRS-----          |
| Cangor  | VPPVASSQAALKLLPQGTSEKVPYGGSHGICSTHKDQINKDIFDISN-----          |
| Sacmal  | VPPAASSKAAIKLLPMPKEKVYKGASHGLCSTHKMQISEDIIYQFVKMS-----        |
| Canaas  | VPIIAASEAAIKLLPKPTVKVCVGGSHGICSTEKDRVNQDIWDFIKE-----          |
| Ogapol  | VPIKISGLKSVKLLQKGTILKVYKGGSHAIHNINVDENVNDLLEFIRG-----         |
| Tricif  | VPFEASGKKSQALLPKQGTILKVYPGGSHAIHNINVDENVNDLLEFISQ-----        |
| Blamok  | VPFEASGKKSQALLSKGSLKVYPGGSHAIHNINVDENVNDLLEFISL-----          |
| Sugxyl  | VPFEISGKKSQALLPKQGTILKVYPGGSHAIHNINVDENVNDLLEFISK-----        |
| Blaser  | VPIETSGHKSQALLPKQGTILKVYPGGSHAIHNINVDENVNDLLEFISQ-----        |
| Exooli  | VPIASAGLKSQALLPKQGTILKVYPGGSHAIHNINVDENVNDLLEFISQ-----        |
| Blaraf  | VPIEASGQKSQALLPKQGTILKVYPGGSHAIHNINVDENVNDLLEFISQ-----        |
| Phicur  | VPPGASGEKSQALLPKQGTILKVYPGGSHAIHNINVDENVNDLLEFIRS-----        |
| Speeur  | VPIGASAHAGVKLAKKGTILKVYPGGSHALPNIAAEVNQDILLDFLQA-----         |
| Molsco  | VPIDDSAREGIKLLKKGTLKVEYPPGSHALPNLNIEEINQDILLDFLKA-----        |
| Hel_sp  | VPIDASAREGVKLLKNGTLKVEYPPGSHALPNIRAEVNQDILLDFLKS-----         |
| Scyliy  | VPIGASAHGEGIKLLKHGTILKVYPGGSHALPNIFTDEVNQDILLDFLKS-----       |
| Cadmal  | VPIAAAAEIGVLLKKGTLKVEYPPGSHALPNMEAQTVNEDLLNDFLKD-----         |
| Cad_sp  | VPIGAAELGVKLLKKGTLKVEYPPGSHALPNMEAQSVNEDLLNDFLKA-----         |
| Claban  | VPIEASALEAIKLLKHGKLVVVGKSHALPNINAEVNAELLFLEEK-----            |
| Clapsa  | VPIEASALEAIKLLKHGKLVVVGKSHALPNINAEELSAELLFLEEK-----           |
| Fonrub  | VPIEASAEAIKLLKNGKLVVEIKGSHALPNINAEVNAELLFLEEE-----            |
| Fonped  | VPIEASAEAIKLLKNGKLVVEIKGSHALPNINAEVNTLELLFLEE-----            |
| Fonmon  | VPIEASAEAIKLLKNGKLVVEIKGSHALPNINAEVNTLELLFLEE-----            |
| Claimm  | VPIEASAEAIKLLKNGKLVVVGKSHALPNIKAEVNSELLKFLEEK-----            |
| Fonere  | VPIDASAEAIKLLKNGKLVVVGKSHALPNIAAEVNNEELLFLEAK-----            |
| Rhilig  | VPIGASAREGILPKGTILKVEYPPGSHALPNIHIEEINQDILLDFLRS-----         |
| Mycgal  | VPIKDAALKGIELLPKGTILKVYLGSHALPNICVDEVNKDLWEFLNA-----          |
| Mycven  | VPIKDAALKGIRLLPKGTILKVYLGSHALPNICVDEVNKDLWEFLNA-----          |
| Neodit  | VPIQASALKAAKVLPKATILKVYPGGSHALPNIHDEVNKDLLEFLKA-----          |
| Exoaqu  | VPIETSGRRVQCKKGTILKVEYPPGSHALPTICDEVNQELLKFLOS-----           |
| Fonmul  | VPIDAGRRVQCKKGTILKVEYPPGSHALPTICDEVNQELLKFLOS-----            |
| Cypeur  | VPIDATGRRSAKLAKKGTILKVEYPPGSHALPTTRINEVNQALLEFLQS-----        |
| Polcit  | VPIDASARKTILKLLKGTILKELPGSHALPTICDEVNEELLFILRS-----           |
| Phiatt  | VPIDTSARRAIKCKTGTILKVEYPPGSHALPSICAEVNQALLEFLQS-----          |
| Lacocc  | VPIAVSAYKTVNLLPDGKLVYQGGSHALPNIFMDELNQDILLDFVRS-----          |
| Pseflu  | VPIEASGIASAALVKGSTILKIYSGAPHGLTDTTHKDQLNADLLAFIKG-----        |
| Syn_sp  | VPIGASALLSAKLVKNSILKIYPPGSHSLGDSMSREQLNADLLFVKS-----          |
| Metace  | VPIGASAMLSKLVKDATLKVIYKGAPHGMCTTHKDQVNEDLLAFFKA-----          |
| Metsil  | VPIGASAMLSKLVKDATLKVIYKGAPHGMCTTLKDRVNEDLLAFFKE-----          |
| Methor  | VPIGASALLSSKLIKSTILKIYKGAPHGMCTTLKDRVNEDLLAFFQGVPRNNKGGEEN--  |
| Met_sp  | VPIGDSALLSSKIVKGAIVKVYKGAPHGMCTTLKQVNEDLLAFFKA-----           |
| Pedpar  | VPIGAAAMQSSKLVKNAVILKVYKGAPHGMCTTLKDRVNEDLLAFFKG-----         |
| Metthe  | VPIGVSAAMLSKIIIRGAILKVYPGAPHGICSTHKDQINADLLAFIKGEAVPDTSMESGMK |
| Wicsor  | VPFEASGKQTVKLIKGAQLISYPGFPHGMPVTHADAINKDILKFIKS-----          |
| Novros  | VPFATTGKLAAQIVPGAQLKSYPGYPHGMPTVTHADQINADLLAFIQQ-----         |
| Wicdom  | VPYKNAAIRQKELLKNGTLKVEYPPGSHGMHTVHADIIINQELLKFRQ-----         |
| Rhoery  | VPIANSATAVTLVKNARKLVYPGLSHGMCTVNADTVNADLLSFIES-----           |
| Burpyr  | VPIADAALKSIKLLQNGTLKTYPGYSHGMLTVNADVLNADLLAFVQA-----          |
| Wicocc  | VPYKLASVLQDKLIRNSTLKVIYKGLSHGMHTTHASDINADILKFIQ-----          |
| Strliv  | VPYADAAPKSAELLANATLKSYEGLPHGMLSTHPEVLNPDLLAFVKS-----          |
| Starrat | VPYHDSGPLSAKLAPKQILITYGAPHGIPTHASKLNEDLLKFVKQ-----            |
| Burcen  | VPYADSGVLSAKLVKNGKILITYGAPHGIPTHADKVNADLLLEFLQS-----          |
| Pseput  | VPYENSGLVLSAKLIPNGALKTYKGYPHGMPTTHADVINADLLAFIRS-----         |
| Sta_sp  | VPFEVSGQKSQALKAPNATLKTYPGFPHGMLTTQADVINAADLLSFIRA-----        |

|         |        |
|---------|--------|
| Debhan  | -----  |
| Meygui  | -----  |
| Cancar  | -----  |
| Cantam  | -----  |
| Pricar  | -----  |
| Canasc  | -----  |
| Wicano  | -----  |
| Canint  | -----  |
| Canbla  | -----  |
| Canaur  | -----  |
| Canhae  | -----  |
| Sacfod  | -----  |
| Hypbur  | -----  |
| Debmar  | -----  |
| Hyphei  | -----  |
| Cangot  | -----  |
| Serodo  | -----  |
| Ser_sp  | -----  |
| Hyphom  | -----  |
| Cangor  | -----  |
| Sacmal  | -----  |
| Canaas  | -----  |
| Ogapol  | -----  |
| Tricif  | -----  |
| Blamok  | -----  |
| Sugxyl  | -----  |
| Blaser  | -----  |
| Exooli  | -----  |
| Blaraf  | -----  |
| Phicur  | -----  |
| Speeur  | -----  |
| Molsco  | -----  |
| Hel_sp  | -----  |
| Scylig  | -----  |
| Cadmal  | -----  |
| Cad_sp  | -----  |
| Claban  | -----  |
| Clapsa  | -----  |
| Fonnub  | -----  |
| Fonped  | -----  |
| Fonmon  | -----  |
| Claimm  | -----  |
| Fonere  | -----  |
| Rhilig  | -----  |
| Mycgal  | -----  |
| Mycven  | -----  |
| Neodit  | -----  |
| Exoaqu  | -----  |
| Fonmul  | -----  |
| Cypeur  | -----  |
| Polcit  | -----  |
| Phiatt  | -----  |
| Lacocc  | -----  |
| Pseflu  | -----  |
| Syn_sp  | -----  |
| Metace  | -----  |
| Metsil  | -----  |
| Methor  | -----  |
| Met_sp  | -----  |
| Pedpar  | -----  |
| Metthe  | PAAGAR |
| Wicsor  | -----  |
| Novros  | -----  |
| Wicdom  | -----  |
| Rhoery  | -----  |
| Burpyr  | -----  |
| Wicocc  | -----  |
| Strliv  | -----  |
| Starrat | -----  |
| Burcen  | -----  |
| Pseput  | -----  |
| Sta_sp  | -----  |